

Inconceivable Minds

Amber Ross

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2013

Approved by:

William G. Lycan

Daniel C. Dennett

Ram Neta

Larry Bryce Huebner

Dorit Bar-On

© 2013
Amber Ross
ALL RIGHTS RESERVED

Abstract

AMBER ROSS: Inconceivable Minds
(Under the direction of William G. Lycan and Daniel C. Dennett)

The current debate over the metaphysical nature of the mind is dominated by two major philosophical views: property dualism and physicalism. According to property dualism, mental properties are of metaphysically distinct types. There are “phenomenal properties”, or “qualia”, the properties that constitute conscious experience, and there are the cognitive or functional properties of the mind. According to physicalism, there is one metaphysical type of mental property, though there may be a deep conceptual divide between experience and cognition.

I challenge both these assumptions. Focusing on two of the most popular anti-physicalist arguments- the Conceivability Argument and the Knowledge Argument- I argue that the property dualist’s account of the relationship between consciousness and behavior- the “explanatory irrelevance” of consciousness to sufficient explanations our behavior and how we make judgment about our conscious experience- make the first argument untenable and the second irrelevant to the metaphysical debate.

I also present a case against a physicalist conceptual separation between cognition and consciousness by challenging the “Phenomenal Concept Strategy”, the most popular argument supporting such a separation. This conceptual separation supposedly allows us to conceive of “philosophical zombies”, creature physically identical to human beings but who

lack conscious experience. The phenomenal concept strategy aims to explain how we can conceive of zombies while maintaining a physicalist account of the metaphysics of mind.

For this strategy to succeed, the physicalist must show that we share our epistemic situation regarding consciousness with our “zombie-twins”. Zombies make claims about their own phenomenal experience, just as we do, but by definition they have none. I examine the most common physicalist interpretation of the zombies’ beliefs about their own conscious experiences and show that this leads to the creation of “inconceivable minds”- creatures whose mental features would be incompatible with the very interpretation of zombie “phenomenal” belief on which this strategy is based.

My dissertation has two overarching goals. First, to undermine the plausibility of the two most popular arguments for property dualism, and second, to force physicalists to reconsider both the phenomenal concept strategy and their commitment to the genuine conceivability of zombies.

Table Of Contents

Introduction- On The Conceivability of Zombies.....	1
I. The Zombie: An Inconceivable Mind?	1
II. The Conceivability Argument Against Physicalism	6
II.1 The Role of the “Explanatory Gap”	12
III. The Property Dualist’s Essential Task	13
III.1 Establishing Epistemic Inequality	14
III.1a Making the Distinction Make a Difference	14
III.1b The Perils of Sharing our “Epistemic Situation” with our Zombie-Twins.....	15
IV. Eschewing Intuition	17
V. Thus Spoke the Zombie	22
1. A Little Background	26
1.1 Qualia: Inverted and Otherwise	26
1.1a Irreducible, Enigmatic, Problematic.....	30
1.2 The Metaphysical Clubs	35
1.2a Property Dualism- Getting Metaphysical Mileage out of Epistemic Evidence	35
1.2a1 We Can’t Tell You Why... Returning to the “Gap”	36
1.2a2 Mind-Splitting- Phenomenal and Cognitive Mental Properties.....	38
1.2b The Physicalists	41
1.2b1 “A Priori” Physicalism: A Straightforward Rejection of “Common” Sense	42
1.2b2 “A Posteriori” Physicalism: The “Common” Sense- Grant Conceivability, Deny Possibility.....	44

1.3 Can the Metaphysically Impossible be Genuinely Conceivable?.....	45
1.3a A Posteriori Physicalism and Kripke’s “New” Necessity: A Plurality of Possible Worlds	45
1.3b Defending CP: The Modal Minimalist	49
2. Oh, The Modality!.....	54
2.1 The “Varieties” of Zombie-Conceivability, and Previous Unsuccessful Attempts to Undermine Them	54
2.1a Were it only so Easy: “Modal Imagination” and the Unfortunate Irrelevance of the First-Person-Subjective Unimaginability of Zombies to the Tenability of the Conceivability Argument.....	57
2.2 Disentangling the Conceivable from the Possible from the Actual but Incredible	64
2.2a Inconceivable Objects.....	64
2.2a1 Escher's Stairs.....	64
2.2a2 Penrose’s Triangle.....	66
2.2a3 An Incredible Actual Object- The Mobius Strip	69
2.3 Possible, Conceivable, or <i>Merely Intelligible</i> ?.....	70
2.3a “Intelligibility”: a Neglected Piece of the Modal Puzzle	70
2.3a1 Intelligibility, Conceivability, and Conceptual Disagreements	73
3. The Dualist’s Essential Task: Defending our Epistemic Situation from the Zombie-Occupation.....	82
3.1 Zombies, the Knowledge Argument, and the “Phenomenal Concept Strategy”	82
3.1a Using “Conceptual Isolation” to Block Entailment.....	84
3.1b The Phenomenal Concept Strategy (an Informal Formalization).....	86
3.2 The Dualist’s Argument for Epistemic Inequality: a “Dilemma” for the Physicalist.....	88
3.3 Dualists’ and Physicalists’ Differing Conceptions of Phenomenal Concepts	93
3.3a A Fundamental Division- Constitutional and Non-Constitutional Phenomenal Concepts.....	93

3.3b Non-Constitutional Accounts	95
3.3b1 Phenomenal Concepts as Pure Demonstratives	95
3.3b2 Phenomenal Concepts as Impure Demonstratives.....	96
3.3c Constitutional accounts	97
3.4 The Property Dualist’s Restrictions on Phenomenal Concepts.....	100
3.4a Pure Phenomenal Concepts and the “near Cartesian certainty” of our Phenomenal Beliefs	102
3.4b The Limits of the Zombie’s “Phenomenal” Concepts and “Phenomenal” Beliefs, and the “Zombie Intentional Object” Thesis	104
3.4b1 Actual Epistemic Access to Phenomenal Properties.....	105
3.4b2 Potential Epistemic Access to Phenomenal Properties	106
3.4b3 The “Zombie Intentional Object” Thesis.....	108
3.5 The Invasion of the “Schmenomenal”-Zombies and the “Schmenomenal” Concept Strategy.....	108
3.5a The Zombies’ “Schmenomenal” Beliefs.....	113
3.5a1 Schmenomenal Beliefs- True, But Complex	114
3.6 Epistemic Equality for All?	115
3.6a Schmenomenal-Zombies: Our Epistemic Twins or Epistemic Equals?.....	116
3.6a.1 The Marys and The Oscars.....	119
3.7 Three Property Dualist Arguments Against Epistemic Equality	120
3.7a An Illicit Appeal to Functional Disparities.....	120
3.7b “ <i>This</i> ” is Rich- A Second, Less Obvious, Illicit Appeal to Functional Disparities	123
3.7c Does Our “Acquaintance” with our Qualia Justify Epistemic Inequality?	128
3.8 The Epistemic Equality of Zombies- Living to Fight Another Day	132
4. The Schmenomenal-Zombie: Friend or Foe? Re-evaluating the “Schmenomenal” Concept Strategy.....	135

4.1 False Beliefs about False Consciousness	135
4.1a What the Schmenomenal-Zombie Knows.....	137
4.1b Inverted-Schmenomenal-Zombie-Twins.....	139
4.2 Epistemically Equivalent Beliefs, or Merely Epistemically Similar?	143
4.2b Investigating the Cognitive Significance of Schmenomenal beliefs	144
4.3 Zombies, Inverts, and Schmenomenality- Peacefully Coexistent or Secretly Antagonistic?.....	145
4.3a The Four Faces of Mary: Our Cast of Characters.....	146
4.3a1 Ordinary Mary.....	147
4.3a2 Inverted Ordinary Mary.....	149
4.3a3 Schmenomenal-Zombie Mary	150
4.3a4 Schmenomenal-Zombie Inverted Mary	152
4.4 “Mary-Relations”: The Deceptive Intelligibility of “Correspondence” and “Inversion”	153
4.4a Inverted-Twinship	154
4.4b Human / Schmenomenal-Zombie-Twinship.....	155
4.5 The Results are In, and Odd:.....	156
4.5a The Relative Cognitive <i>Insignificance</i> of Inverted Schmenomenal-Zombie Beliefs.....	158
4.6 The Inverted-Schmenomenal-Zombie-twin: a Genuinely Inconceivable Mind?.....	161
4.6a Epistemically-Quite-Similar Beliefs?.....	163
5. The Conceiver Stays in the Picture: Why “Ideal Conceivability” is a Less-Than-Perfectly-Rational Notion.....	165
5.1 Carving up “Conceivabilities” to Explain Away the Appearance of Subjectivity.....	165
5.1a Peering Into the Possible?	166
5.1a1 Yablo’s criticism- Any conceiver could be mistaken.....	166
5.2 Idealizing Conceivability?	168

5.2a The (Less-Than-Perfectly) Rational Notion of “Ideal Conceivability”	171
5.3 IREA: The Ideally Reasonable Epistemic Agent	172
5.4 IREZA: An Ideally Reasonable Epistemic Zombie-Agent.....	176
5.4a Experts without Expertise.....	177
5.5 Conceptual Incompetence or Hallucination? The Undetectable Errors of an Ideally Reasonable Epistemic Agent.....	178
5.5a How do you solve a problem like IREZA?	181
5.6 Conceivability by Fiat	182
5.6a A Double Standard for Intuition? Denying the Conceivability of Physicalism	184
5.6a1 Chalmers’s Response to the “Anti-Zombie” Argument.....	184
5.6b A Broken Epistemic Arm.....	190
6. Mystery or Paradox? “Explanatory Irrelevance” and the Knowledge Argument.....	193
6.1 Phenomenal Belief, Phenomenal Judgment, and Explanatory Irrelevance.....	193
6.1a The Paradox of Ideal Conceivability.....	197
6.2 Hypersensitive Mary and The Knowledge Argument.....	198
6.2a A Priori Physicalism, and the Significance of “Knowing Red on Sight”	203
6.2b Introducing “Hypersensitivity”	205
6.3 Mary’s Story, From the Beginning.....	207
6.3a Behaving Properly- The Orthodox Zombie Mary Story	209
6.4 The Hypersensitive Mary Conceivability Argument	212
6.4a Hypersensitive Mary’s ‘Ho, hum’ Rebellion	213
6.4b Hypersensitive Mary in Nida-Rumelin’s “Color Room”	214
6.5 “Knowing by sight”?	215
6.5a Surprise, or Suppressed Premise?	216

6.5a1 The “Slow and Laborious” Assumption.....	217
6.5a2 The “Wrong Kind of Knowledge” Premise.....	219
6.6 The “Cognitive Stance”	221
7. A Conclusion.....	224
Bibliography.....	226

Introduction- On The Conceivability of Zombies

I. The Zombie: An Inconceivable Mind?

“Zombies” have received more attention in recent years than any other character that philosophers have lifted from popular culture for their own device: more than Davidson’s Swamp Man, Lewis’s Martians, or Putnam’s Brain in the Vat. The zombies of philosophy are fascinating and radically different from their pop-cultural namesakes. As no philosopher fails to note, these are not the kind of zombies that stumble around semi-alert. “Philosophical Zombies” look and act just like ordinary human beings; by all outward appearance zombies and humans are indistinguishable.

Every person has a zombie-twin, a creature behaviorally identical to him or herself. Zombies resemble us in all our physical and cognitive respects; they act as we act, speak as we speak. The only difference between a zombie and a human being is that none of the zombie’s physical activity- inner or outer- is accompanied by any conscious experience. Our zombie-twins’ senses gather all the same information from the world that ours gather, but- unlike ourselves- they have no experience of it whatsoever. On one common meaning of mind- the mind as a “stream of consciousness” or “phenomenal field”- philosophical zombies are literally *absent-minded*.

David Chalmers, who will be our stalking-horse zombie-advocate, describes the philosophical zombie as follows:

A Zombie: someone or something physically identical to me... but lacking conscious experiences altogether... He will certainly be identical to me functionally... he will be psychologically identical to me... It's just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie. (Chalmers, 1996: 94-95)¹

The sole difference between a human being and her philosophical-zombie-twin is that, where the first has a rich inner mental life, her zombie-twin has nothing resembling an inner life.² A philosophical zombie is all brain and no consciousness. Wondering what it would be like to be a philosophical zombie would be an exercise in futility; by definition, there is nothing it's like to be a philosophical zombie. Life as a zombie wouldn't be boring or dull; being a zombie wouldn't be "like" anything at all.³

Though there's nothing it's like to be a zombie, Robert Kirk gives a remarkably florid description of philosophical "zombiehood".

imagine that somewhere in this or another world there is an exact physical double of yourself. It not only looks and behaves like you, it matches you in every detail of body and brain: it is a particle-for-particle duplicate. So we can assume it says and writes exactly the same things as you do. In my own case this creature talks a lot about consciousness... Naturally everyone treats it as if it were conscious. Not only is that attitude natural; it seems to be supported by overwhelming evidence. How could this creature talk and write about consciousness unless it were conscious? But the example is strictly philosophical, and this particular physical duplicate is a

¹ Chalmers uses "psychological" in a very specific sense, which we will discuss further along. In the meantime, "psychological" mental states can be understood as whatever we might think of as a *cognitive* act or state:

² Zombies retain some of our attributes that philosophers would call "mental states": intentional states, propositional attitudes with certain, non-phenomenal, objects; they have all the standard properties of a mind aside from the feeling of conscious experience. On certain colloquial meanings of "mind", this absence is enough to say that zombies are "mindless". After all, the *philosophical zombie* wasn't named after the pop-cultural *Zombie* for nothing.

³ Some philosophers have used the fact that wondering what its like to be a zombie is "an exercise in futility" to dismiss the conceivability of zombies altogether. I think these objections are insightful but miss the mark, and we will discuss some of them shortly.

philosophical zombie. By definition philosophical zombies are supposed to have no conscious experience at all: ‘all is silent and dark within’.⁴ (2005: 3)

How could a creature, one physically identical to yourself, talk and perhaps write about consciousness, describe and answer questions about his or her thoughts and feelings precisely as you would, all the while having never had anything like a conscious experience? This peculiarity of zombiehood is the grounds for what I call the blunt “absurdity objection”. Are we actually supposed to believe that a creature could talk about consciousness, argue over the truth of physicalism and property dualism, describe “its experience” in as much detail as we do, and yet not be conscious itself? The notion is absolutely absurd!⁵ These phenomenally bereft creatures *seem* to make judgments that are about conscious experience, judgments indistinguishable from our own, though they have no conscious experience themselves. My zombie-twin firmly “believes” she is conscious, that she has a rich inner life; she even “believes” she has vivid experiences within her dreams. But according to most philosophers, these “beliefs” are all false. What seems worse, my zombie-twin has *no way of discovering* her enormous “mistake”.

How are we to understand the difference between our zombie-twins’ “beliefs” about what they *think* is their own conscious experience and our beliefs about our *real* conscious experience?⁶ Why should we believe there is any *real* difference between ourselves and the philosophical zombies?

⁴ The words are Iris Murdoch’s in a discussion of behaviorism: *The Sovereignty of the Good*, 13. (footnote appears in original Kirk quote).

⁵ For a compelling “argument from absurdity”, see Dennett (1998), “The Unimagined Preposterousness of Zombies”.

⁶ For this sort of objection to the conceivability of zombies, see Shoemaker (1975) and (1981) “Absent Qualia are Impossible: a Reply to Block”.

Most philosophers are not compelled by the “absurdity objection” and have the strong intuition that these zombies are *conceivable*. Some argue that these mentally-divided creatures are metaphysically *possible*. And in reading the description of zombies and the story of zombie-life above, it seems as though we can imagine the zombie and its zombie world, at least to some degree. The philosophical zombie must be one of the most curious creatures imaginable, if it is imaginable, and one of the most philosophically powerful as well. The mere possibility of this kind of zombie has the power to overthrow what was- until recently- an almost universally accepted position in philosophy of mind: physicalism. Physicalism, in the most general terms, is the view that all existing things, at their most fundamental level, are composed of or reduce to or supervene upon only *one kind of* property- physical properties- the kind of properties that compose everything we think of as ordinary physical objects. If physicalism is true, then any creature physically identical to a human being, living in world relevantly similar to our own, will be identical to a human being in all its respects, including its conscious experience.⁷ Though any philosopher may claim that she or he can conceive of zombies, according to physicalism they are merely one among many conceivable *impossibilities*.

Property dualists who argue for the metaphysical possibility of zombies, as Chalmers does, believe that the physical, functional, and structural properties of our mental states reduce to or supervene upon more fundamental properties of the same metaphysical kind; physical, functional, and structural properties that are more fundamental than those mental

⁷ I add that the world must be similar to ours in relevant respects, since, for example, a creature physically identical to a human being but in a world with the wrong properties may be inert rather than conscious.

states themselves.⁸ But the dualists argue that the fundamental features of our world are not exhausted by physical properties alone. There must be at least one other type of property, a property radically different from the functional, information processing, problem solving, *cognitive* kind of mental property. These properties are “qualia”, the “phenomenal properties” of our mental states, properties that, according to the dualists, are metaphysically different in kind from our psychological or cognitive mental properties. These properties do not reduce to nor do they metaphysically supervene upon any *non*-phenomenal properties. The presence of these properties cannot be explained in terms of physical, functional, or structural properties, and they are the only kind of property that, by hypothesis, you do not share with your philosophical-zombie-twin. Our qualia, our phenomenal properties, are the properties that constitute the character of our conscious experience; having or lacking these properties will determine whether you are conscious, or your mind is like that of the zombie-all silent and dark within.

“Naturalized” property dualism assumes that there are “psychophysical laws” in the actual world, laws that govern the relationship between our cognitive and phenomenal mental states such that, as a matter of empirical fact, conscious experience *nomologically* supervenes upon the physical, functional, and structural properties of our world.⁹ But they argue that the

⁸ For the most part, I will be using the terms ‘physical’, ‘functional’, ‘cognitive’, ‘structural’ and ‘psychological’ interchangeably when describing mental properties that are not phenomenal properties. When the differences between these types of properties are relevant I will make it clear which I specifically intend.

⁹ Loosely, one can describe the supervenience relation in this way: In all cases of supervenience, B-properties supervene upon A-properties if the A-facts (facts about A-properties, the “supervenience base”) in some way necessitate the B-facts (facts about B-properties, the properties that “supervene” on the supervenience base). We could say that A-properties always bring about the B-properties, or that B-facts cannot change without some change in the A facts.

The two kinds of supervenience relevant to this discussion are *natural* (“*nomological*”) and *logical* (or *metaphysical*) supervenience. B-properties *naturally* or *nomologically* supervene upon A-properties when-

“phenomenal facts” of our world- facts about the character of our conscious experience- do not *logically* or *metaphysically* supervene upon physical facts. If phenomenal facts logically supervene upon physical facts, it would be logically necessary that the physical facts of a world *entail* the phenomenal facts. To put the point another way: for phenomenal facts to logically or metaphysically supervene upon physical facts, it would have to be logically *impossible*, or conceptually incoherent, or inconceivable, for the character of one’s conscious experience to change without there being some change in the physical properties of the world.¹⁰ Biological facts, for example, logically supervene on physical facts. We could say, following Chalmers (who borrows an image from Kripke, (1972)), that once God fixed all the physical facts of the world, the biological facts came along for free. But after setting the physical facts in place, God had “more work to do to make sure there is a law relating the [physical] facts and the [phenomenal] facts” (1996: 38). According to property dualism, biological, chemical, and *psychological* facts logically supervene upon physical facts; phenomenal facts stand apart as metaphysically and conceptually independent of any other kind of fact.

because of the particular natural laws governing our world- facts about B are determined by (or “fixed”) by facts about A. For example, assume (whether or not this is true) that a perpetual motion machine cannot exist in our world. That is to say, given the actual laws governing our world, a perpetual motion machine cannot actual exist in it. So a perpetual motion machine would be naturally or nomologically impossible. If we consider facts about perpetual motion machines to be “B-facts”, and facts about the natural laws of our world the “A-facts”, then the fact that a perpetual motion machine cannot exist in our world *nomologically supervenes upon* the facts about the laws governing our world. The impossibility of a perpetual motion machine nomologically supervenes upon the laws of our world. In order for a perpetual motion machine to be *possible in our world* (in order for the B-facts to change) there must be some *change in the natural laws* of our world (the A-facts must change).

¹⁰ The terms “metaphysically supervene” and “logically supervene” are used synonymously in the literature and I will be using them this way from here forward. The terms “conceptually incoherent” and “inconceivable” are likewise synonymous in the literature we will be discussing, and I will be using these interchangeably as well.

II. The Conceivability Argument Against Physicalism

Assume for a moment that the dualist is correct, and that facts about the character of our conscious experience are not logically determined or entailed by facts about our physical nature. If this is logically possible, then it is logically possible or conceivable for a subject's physical properties to be exactly as they are now while her conscious experience differs in some way. It would be conceivable for you to feel a slight ache in your left foot that you are not actually feeling now though your physical body- including your brain activity- is in a state that is molecule-for-molecule identical to its current state. That *ache* would be a phenomenal property of your experience, and if it is conceivable for there to be a small change such as this in your phenomenal properties while your physical properties remain the same, the door of conceivability swings wide open for the conceivability of philosophical zombies. If physical facts do not logically entail phenomenal facts, then it is conceivable for all the physical properties of the world to be just as they now though the phenomenal properties differ in some respect. And if it is conceivable for phenomenal properties to change or differ without a change or difference in physical properties, it is conceivable that phenomenal properties could be *absent* from a world altogether though the physical properties of that world are identical to those of the actual world which, according to the dualist, contains both physical and phenomenal properties.

Worlds that are physically identical to our own but lack phenomenal properties are "zombie-worlds", containing "zombie-minds", and it follows from the conceivability of the less-radical scenarios above that zombie-worlds and zombie-minds conceivable as well. According to property dualists who argue as Chalmers does, so long as zombies are

genuinely conceivable they are *possible*.¹¹ And if zombies are possible, physicalism is false.

If zombies are possible, then it is possible that the phenomenal consciousness we find present in our world- the feature that separates our world from zombie-worlds- could have been *absent*, while the physical facts remained just as they are now.

But this is not how our world turned out; our world does contain phenomenal consciousness. We have conscious experience. If our world is not a zombie-world, our world must contain an additional kind of property from those that exist in zombie-worlds; it must contain *phenomenal* properties, or “qualia”, as well.

The conceivability argument may be the most compelling for property dualism, and we will generally refer to the premises of the conceivability argument in this simple form:

1. Zombies are conceivable
2. Whatever is conceivable is possible
3. Zombies are possible
4. Physicalism is false

But it can be laid out more technically as follows:

1. $P \& \sim Q$ is conceivable.
 2. If $P \& \sim Q$ is conceivable, $P \& \sim Q$ is metaphysically possible.
 3. If $P \& \sim Q$ is metaphysically possible, materialism is false.
 4. Materialism is false.
- (Chalmers, 2010: 142)

As Chalmers lays out the conceivability argument here, P is “the conjunction of all microphysical truths about the universe”, and Q is “an arbitrary phenomenal truth” (ibid: 142). A “phenomenal truth” is a particular fact about conscious experience; it can be a fact about a particular individual’s conscious experience or about phenomenal consciousness in

¹¹ We will discuss the reasoning behind this claim- the notion that conceivability entails possibility, or “the conceivability/possibility principle”, in much more depth as we go forward.

general. So premise 1 of the conceivability argument states that you can conceive of a world in which the phenomenal facts are not necessitated by the physical facts; in other words, it is conceivable that the “psychophysical conditional” is false. And most physicalists accept premise 1; they agree that we can conceive of the physical facts of our world being just as they now and the phenomenal facts differing in some way.

If you are among the majority of philosophers, you also believe you can conceive of a world physically identical to our own in which your physically-identical twin might have a conscious experience as of perceiving red while you have a conscious experience as of perceiving green. That is, you believe you can conceive of having an “inverted-twin”. And you believe you can conceive of a creature molecularly identical to yourself, living in a world molecularly identical to our own, who is entirely void of conscious experience: your zombie-twin.

Premise 2 of the conceivability argument is what Chalmers and others have called the conceivability/possibility principle, or “CP” for short. According to CP, whatever is genuinely conceivable is possible. If zombies are genuinely conceivable, then zombies are possible, and if zombies are possible, physicalism must be false. Of course, there are many uses of the term “conceivable”, and not types of conceivability will entail possibility. As Chalmers distinguishes between uses of the term, there are three important dimensions along which the term can be divided, creating at least 9 distinct “varieties”, the details of which will be discussed in due course. The importance of distinguishing between these “varieties of conceivability” is that it allows Chalmers to separate and *isolate* the different ways in which the term is used. This separation makes it more plausible that there could be at least one “variety” of conceivability which is a purely “rational notion”, something akin to the notions

of a priority and entailment. The crucial feature of this “variety” of conceivability is that it can be determined objectively; if this is correct, then there will be at least one kind of conceivability-judgment that is not relative to the beliefs, and epistemic frailties, of the conceiver.

The gold standard for conceivability, the kind that is synonymous with “conceptual coherence”, is what Chalmers calls “ideal conceivability”. Judging whether a scenario is *ideally conceivable* requires scrutinizing the contents of the concepts composing that scenario and the logical relations between those concepts. A scenario is ideally conceivable if there is no contradiction within it, and it will be ideally *positively* conceivable if a subject could, in principle, “imagine” a world that fits the scenario’s description and “fill in” arbitrary details of that scenario without uncovering any latent conceptual incoherence.¹² Property dualists claim that zombies are ideally conceivable, that the zombie-scenario is conceptually coherent, and from this objective, “rational”, notion of conceivability Chalmers infers their possibility.

All parties assume that our world contains “psychophysical” laws governing the connection between our cognitive and phenomenal states, nomologically guaranteeing that no one in *this* world is a zombie of the philosophical sort. But if zombies are possible, then it is possible that in some other world- one with different laws governing mental and physical properties- i.e., a world in which the link between *physicality* and *mentality* is broken or nonexistent- creatures that seem like ordinary human-type beings are unconscious zombies instead. But according to physicalism, every creature, in every possible world, that is

¹² This sort of imagination intended here is “modal imagination”, distinct from perceptual imagination; we will examine this type of “imagining” more closely in chapter 2.

molecule-for-molecule physically identical to a human will have human mental properties, both cognitive and phenomenal.¹³ This means that the mere *possibility* of a zombie-world makes physicalism an untenable position; physicalism would be false in our world and in all possible worlds. If zombies are possible, physicalism is *necessarily* false.

Physicalism is not merely a mainstream philosophical view but a fundamental assumption in the scientific study of the mind. If physicalism is false, then our minds must be more than brains-in-action. Our conscious experience may be nomologically connected to our cognitive, functional, and structural states as a matter of empirical fact, but as we said above, our world *could have turned out differently*. If our world could have turned out differently- *physically* just the same but *phenomenally* different- there must be some additional feature of our world that *could* have been absent from it but is in fact present. Perhaps this feature “emerges” from our brain activity, but nevertheless it is a fundamentally different kind of property than our physical properties, and it could only be this additional, non-physical, property that accounts for our conscious experience.

Scientific explanation, in its current form, does not have room for a non-functional, non-physical, non-structural fundamental property. A basic assumption of empirical science is that the phenomena of our world can be explained without appealing to anything beyond the structure of their physical properties and the relations between them.¹⁴ If zombies are possible then our current form of scientific explanation, one that- in principle- can provide a

¹³ This way of putting the point assumes that there is a real difference between the “two” kinds of mental properties.

¹⁴ Here I am envisioning forces and fields as fundamentally physically, something like structure or relations between physical properties. All that really matters here is that we need not posit an additional metaphysical category to explain their existence, and even if we do think our best scientific explanation of these features considers them metaphysically different from “ordinary” physical properties, for now I am only concerned with mental properties- related kinds of dualism are fascinating but outside the scope of this discussion.

sufficient explanation for every other fundamental aspect of our world, could not explain conscious experience. If property dualism is true, the mind stands outside the scope of empirical investigation.

Zombies are so philosophically powerful that their mere conceivability, if this entails their possibility, ensures that science as-we-know-it will never explain the conscious mind.

II.1 The Role of the “Explanatory Gap”

If it is impossible *in principle* for our current form of scientific explanation to account for conscious experience, there will be a persistent and unbridgeable “explanatory gap” between truths about the physical world and truths about the conscious mind. The role of the zombie is to hold this gap open wide enough to make property dualism an intuitively plausible theory of mind. And the more compelling their conceivability, the harder it is for physicalism to *close* the gap- that is, to give a “satisfying” explanation of conscious experience in physical terms- or even to dismiss the gap as insignificant, perhaps arising from some feature of our *concept* of conscious experience and our concept of the mind but irrelevant to the *metaphysical* nature of consciousness itself. The dualist takes the persistence of the explanatory gap- the failure of physicalism to give a “transparent” explanation of the connection between facts about consciousness and physical facts- as conclusive evidence for the falsity of physicalism.

Unlike the formalized conceivability argument above, arguments for property dualism based on the existence of the explanatory gap must rely on two somewhat contentious assumptions. The less contentious assumption is that the explanatory gap between physical and phenomenal facts is permanent. Only a small set of physicalists deny this first

assumption. The majority of physicalists- those who agree that zombies are conceivable- will accept the first but deny the second assumption, the assumption that this unbridgeable gap is *conclusive* evidence against physicalism.¹⁵ Physicalists hold that the conceivability of zombies and the existence of the explanatory gap only provide *prima facie* support for the claim that there is an ontological rift between physical and phenomenal mental properties; the evidence that zombies and the explanatory gap proved for this ontological separation between the physical and phenomenal is still *defeasible*. The physicalist might “explain away” this *prima facie* evidence for property dualism by showing that the conceivability of zombies and existence of the explanatory gap are merely intriguing but ultimately benign consequences of the way our mind organizes and draws connections between our *concepts* of consciousness *and* physicality. Such an alternative explanation would prohibit the dualist from making a direct inference to the possibility of zombies, and the falsity of physicalism, from their conceivability and the existence of the explanatory gap.

III. The Property Dualist’s Essential Task

Showing that zombies are conceivable will ruffle very few physicalist feathers unless the property dualist can accomplish a further task. Any property dualist- whether she appeals to the conceivability/possibility principle to argue against materialism or finds support for property dualism on independent grounds- must establish that there is a relevant difference between *our* epistemic situation regarding consciousness and the explanatory gap, on the one hand, and the *zombies*’ epistemic situation regarding the explanatory gap between

¹⁵ The small set of physicalists we will call “a priori physicalists”, and the majority of physicalists “a posteriori physicalists”, for reasons to be explained later.

“phenomenal” and physical facts that they believe exists in their zombie-world, on the other. The property dualist uses epistemic evidence to ground their metaphysical claims, and if the dualist cannot establish a relevant difference between our epistemic situation and that of our zombie-twins, our own epistemic situation gives us no reason to believe property dualism is true of our world while false in the zombie-world.

III.1 Establishing Epistemic Inequality

III.1a Making the Distinction Make a Difference

As your physical and functional duplicate, your zombie-twin will have a zombie-state identical to each of your *non-phenomenal* states: your cognitive states, psychological states, functional, physical, relational, even *emotional* states insofar as those states have cognitive in addition to phenomenal components. He will “think” and hold “beliefs” about the metaphysics of mind. He will read papers and have conversations about “consciousness”. He will have opinions or “beliefs” about his own mental states, and unless he is an eliminativist, he also “believes” he is conscious. He probably “believes” that there is an explanatory gap between physical and “phenomenal” facts of his world, and “believes” he can conceive of zombies.

Chalmers (2007) defines “sharing an epistemic situation” as “have[ing] corresponding beliefs, all of which have corresponding truth-value and epistemic status, as equally justified or unjustified, cognitively significant or insignificant” (177), and argues that we do *not* share our epistemic situation with our zombie-twins.

Why think zombies do not share our epistemic situation?... On the face of it, zombies have a much less accurate self-conception than conscious beings do. I believe that I am conscious, that I have states with remarkable qualitative character available to introspection, that these states resist transparent reductive explanation, and so on. My

zombie twin has corresponding beliefs. It is not straightforward to determine just what content these beliefs might possess. But there is a strong intuition that these beliefs are false, or at least that they are less justified than my beliefs. (2007: 177)

On this interpretation of “epistemic situation”, for our situation to differ from that of our zombie-twins’, our phenomenal beliefs must differ from our zombie-twins’ corresponding “phenomenal beliefs” in either their truth-value, epistemic status, or both. The question of the *content* of the zombies’ “phenomenal” beliefs is left open, but insofar as these beliefs correspond to their human-twins’ phenomenal beliefs, zombie-beliefs must be either false or less justified.¹⁶

As for our beliefs that do *not* involve phenomenal experience, it seems that zombies and humans may share their epistemic situation. For example, we could share our epistemic situation regarding the “primary” properties of things in the world with our functionally identical zombie counterparts. My zombie-twin and I may both know that the skin of the apple is smooth, but only I know *what it's like* to have the experience of feeling that smooth skin on my hands. Knowledge of the apple’s surface properties is an epistemic situation that we might share in common; knowledge of *what it's like* to have the conscious experience of that surface property is not.

III.1b The Perils of Sharing our “Epistemic Situation” with our Zombie-Twins.

Chalmers claims that there can be no fully physical explanation of the “key psychological features” of our epistemic situation- by which he means the features of our

¹⁶ There is at least one other option: it is possible that the only interpretation of zombie-phenomenal talk that is consistent with either property dualism or a posteriori physicalism is one according to which the zombie’s utterances are meaningless. In Ross ((2013), in progress) I am exploring this third option, but we will not be discussing it in depth here.

phenomenal concepts that supposedly give rise to the explanatory gap and allow us to conceive of zombies- because any physical account of these “key features” will assign our epistemic situation to the zombies as well as ourselves, and, as he claims, the zombies do *not* share our epistemic situation. Our situation is unique in that we are “confronted with...distinctive epistemic gaps” between physical and phenomenal facts (2007: 172). The physicalist seems entitled to question this assumption: perhaps there is a way of interpreting our epistemic situation such that we *do* share our epistemic situation with our zombie-twins. If the physicalist can make such a case, it would be tantamount to giving a fully physical account of our epistemic situation. Showing that zombies *are* our epistemic equals, that we are in the same epistemic situation as our phenomenally-bereft twins, would show that- in principle- there is a fully physical account of our epistemic situation to be given, even if we cannot find it *in practice*. As such, the physicalist would safeguard the tenability of physicalism as a theory of mind.¹⁷ Chalmers concurs on this point,

[E]ven if consciousness cannot be physically explained, we might be able to physically explain [the psychological features of our mind that create our] epistemic situation... If we could physically explain why we are in such an epistemic situation, we would have done the crucial work in physically explaining the existence of an explanatory gap. (2007: 175)

Zombies, our phenomenally-deprived physical duplicates, share all and only our physical properties, so an explanation of why we are in the same epistemic situation as these creatures would be an explanation of our epistemic situation that nowhere needs to posit non-physical properties. It would suffice as an explanation of why there is an epistemic gap

¹⁷ This would be true even if, as a posteriori physicalists claim, the particular nature of our phenomenal concepts keeps us from giving an account of phenomenal consciousness *itself* in fully functional terms.

between facts about the mind presented in phenomenal terms and facts about the mind presented in entirely physical terms.

This “task” reveals further details about the philosophical zombie. Originally, he was defined as your molecule-for-molecule physically-identical but phenomenally-bereft twin. Perhaps implicit in that definition is the qualification that his “phenomenal” beliefs must be epistemically inferior to yours, or perhaps this is an addendum to the original definition of the philosophical zombie. But it is clear that, for the zombie’s conceivability to pull any philosophical weight, the epistemic status of his “beliefs”- at least his beliefs about what he calls “consciousness”- must be inferior to yours. For your zombie-twin to be a “*proper* philosophical zombie”, his “phenomenal beliefs”, whatever content we assign them, must be either false or less justified than your phenomenal beliefs. If we interpret our zombie-twins’ “phenomenal beliefs” as both true and as well justified as our corresponding phenomenal beliefs, the zombie’s conceivability will be irrelevant to any argument for property dualism.

IV. Eschewing Intuition

Intuitions are the invaluable instigators of philosophical inquiry. Our intuitions tell us that there is some notion in the vicinity- be it consciousness, or morality, or knowledge-worth thinking about. But precision- conceptual analysis- is not intuition’s specialty; it is not suited for such “heavy lifting”. In order to give the zombies the serious consideration they are due, we need to extend our sight past the inclinations of our bare intuitions. Often the conceivability argument is treated as an argument from intuition, from a “zombic-hunch”, and the justification for the claim that zombies are conceivable often appears to come from intuition alone.

In this context, we need to reject the notion that conceivability-claims, claims about the conceivability of a particular scenario- be it a twin-earth scenario or zombie-scenario- can be justified by intuition, even popular intuition. Conceivability, as it is used in the conceivability argument, is synonymous with “conceptual coherence” and “conceptual consistency”, and claims about the coherence or consistency of a scenario cannot be justified by appeal to intuition alone. Intuition can lend support to a claim but does not itself constitute an argument for one, even for a claim about conceivability. It can only gesture toward an answer. The faculty of reason must tell us what it is that our intuition points us toward, and how we ought to think about it. Yet the heaviest weapon in the philosopher’s arsenal sits squarely on intuition’s shoulders: *logical possibility*. Chalmers’s informal “argument” here illustrates this tendency to justify our claims about conceivability by appeal to intuition.

I confess that the logical possibility of zombies seems equally obvious to me [as that of a mile-high unicycle]. A zombie is just something physically identical to me, but which has no conscious experience—all is dark inside. While this is probably empirically impossible, it certainly seems that a coherent situation is described; I can discern no contradiction in the description. In some ways an assertion of this logical possibility comes down to a brute intuition, but no more so than with the unicycle. Almost everybody, it seems to me, is capable of conceiving of this possibility. (Chalmers 1996: 96)

There are several related difficulties with this passage. The purpose of drawing an analogy between the intuition that a mile-high unicycle is logically possible and the intuition that zombies are logically possible is to make *justification by brute intuition* in the case zombie-conceivability case look *as appropriate* as it would be for the claim that mile-high unicycles are logically possible. But there is no analogy between the assertion that a mile-high unicycle is a logical possibility and the assertion that zombies are a logical possibility that has anything to do with *justification by intuition*.

The conceivability of a mile high unicycle does *not* come down to a “brute intuition”; The properties that constitute “being a mile-high unicycle”- unlike the property that separate us from our zombie-twins- *do* reduce to more basic, mostly physical, properties, and the concepts that denote these properties- unicycle, mile, height, etc.- are far from controversial. The nomological possibility that a mile-high unicycle could exist depends on a smattering of empirical factors, e.g., physical laws, gravity, the compressive and tensile strength of materials out of which a unicycle could be built. Given our current empirical knowledge we could likely determine whether a mile-high unicycle could exist *in reality*. But more importantly, we can determine whether a mile-high unicycle could exist *in principle* by referring to the uncontested content of the concepts *unicycle*, *mile* and *height*.

Unless my knowledge of these concepts is radically deficient, it seems clear that our belief that a mile-high unicycle is logically possible is justified by analysis of the concepts alone. There is nothing about the claim, “a mile-high unicycle is logically possible” that either “comes down to,” or rests on *brute intuitions*. Its logical possibility is guaranteed by the *stipulated* content of the relevant concepts, which are stipulated by a community of language users. There is no room for appeal to intuition in the unicycle case- the stipulated content of the concepts *unicycle* and *mile-high* give an unambiguous answer- my prima facie judgment is that the answer is “yes”. The only potential controversy would concern empirical questions, perhaps, for example, how its height should be measured.¹⁸

¹⁸ Any potential *conceptual* disagreements that can be resolved by the “agreement to disagree” or a further analysis of related (ordinary) concepts, such as whether- in order for something to qualify as a *unicycle*- there must be some *actual* subject capable of using it for locomotion. In such a case, the question is not whether the object *itself* is possible but whether unicycles are necessarily means of locomotion. And this *apparent* conceptual disagreement may still be resolved by appeal to empirical facts about our world, facts about whether there is such an agent (or perhaps very large groups of agents, such as the nation of China).

I will insist- and I think this is uncontroversial- that “brute intuition” simply cannot justify any premise in an argument for a complex and highly technical metaphysical position. In such arguments, there is always room for debate over the content of the concepts themselves, *even if everyone capable of reasoning about the subject believes they share the same intuition about that content.*¹⁹

This specious “unicycle” case of justification-by-intuition fails to make *brute intuition* look like *reasonable justification* for the claim that zombies are logically possible. At the end of the day, the grounds for a claim that a particular fundamental metaphysical view of the mind *may* come down to nothing more than brute intuition. But it is as clear as the logical possibility of a mile-high unicycle that “brute intuitions” cannot *settle* such an issue. And though intuition may be the *actual* grounds for one’s claim that a particular fundamental theory of mind is superior to another, intuition does not justify such a claim.

Daniel Dennett has said that “[r]igorous arguments only work on well-defined materials,” and because his aim in disputing the property dualist has often been to “destroy our faith in the pretheoretical or “intuitive” concept” of qualia (or “phenomenal” consciousness), his method in dealing with zombies has employed what he calls “intuition pumps”, pumping our intuitions *away* from their conceivability (1990, 251). And here he is not wrong: treating the zombie-conceivability argument as an argument from intuition has led and will continue to lead to nothing but philosophical wheel-spinning; each physicalist

¹⁹ If someone were to claim that they had discovered a truly irrefutable argument, and that at least one of the premises was justified by brute intuition, I honestly would not know what they meant, and would think they misunderstood some concept or other- perhaps *irrefutability*, *justification*, *intuition*, or even *argument*.

objection and dualist reply supported by fundamental and incompatible “brute intuitions”.²⁰ Rather than taking aim at a zombie *intuition*, we need a thorough analysis of the zombie concept itself: an explicit account of the overt and covert features of their “minds”, presented in terms that can be understood by all parties to the debate. We can analyze a zombie-*concept* for coherence, the zombie-scenario for consistency, and the zombie argument for strength; we cannot perform this treatment on a brute intuition. Only after elaborating and analyzing a more precise account of the zombie-concept can we definitively determine whether the zombie is actually conceivable or merely a robust *figment of philosophers’ imaginations*.

The intuition that zombies are conceivable is hard to suppress, but to make a case against the conceivability argument, either dismissing it as ineffectual or denying the conceivability of zombies altogether, we need not explain *away* the forceful intuition or even why we *have* the intuition. If our analysis of the coherence of the zombie-concept clashes with our zombie “intuitions” we should not think of this as a *failure* of intuition. Intuition has done its duty- it has guided us toward a philosophical puzzle. I suggest that we acknowledge- then bracket- our intuitions regarding this metaphysical minefield. Intuition has been called upon enough in philosophy of mind; it is weary of the work. If “intuition” itself becomes a “well-defined” notion, one with which we may critically engage, we may be able to call on it in “rigorous argument”. Otherwise, the task of finding the root of the problem and solving this philosophical puzzle should be left in the hands of Reason.

²⁰ As we will discuss later, treating the pronouncements of intuition as authoritative in justifying conceivability-claims can undermine the zombie conceivability argument altogether.

V. Thus Spoke the Zombie

There has always been a problem of “making sense” of what the zombies are doing when they talk about phenomenal consciousness. They seem to express beliefs about their own conscious states, about the conscious states of others, and about the content of their phenomenal concepts. The importance of this peculiarity- the zombies’ “phenomenal” talk- could be easily overlooked. In focusing on the intended role of the zombie- i.e., as a thought-experiment aimed at understanding the metaphysical nature of consciousness- zombie *concepts* may slip by unnoticed. We have had more pressing questions to answer, and seriously analyzing zombie-talk may have seemed too pedantic, not relevant to the overall aim of the debate. As Chalmers describes our approach to thought-experiments in conceivability arguments,

A typical philosophical thought-experiment starts with *prima facie* positive conceivability. A subject does not imagine a situation in fine detail: microphysical details are usually left unspecified. (2002: 153)

Admittedly, when we are making *prima facie* judgments about conceivability, this is precisely the sort of detail we are licensed to overlook. But issues concerning the similarity and difference between our “epistemic situation” regarding consciousness and the zombie’s corresponding situation have brought zombie-“phenomenal” concepts to the fore, making an examination of zombie-“phenomenal” concepts all the more urgent. If we are trying to determine whether the zombie-scenario is *ideally positively conceivable*, conceivable in the way that is relevant to the conceivability argument, we no longer have license to overlook these “fine details”.

Chalmers (1996) acknowledges the peculiarity of zombie “phenomenal” talk, that one might wonder whether zombies have phenomenal concepts that could refer or had meaning and what might make us answer one way or the other.

One might be worried by the fact that the concept of consciousness is arguably not present [in] the zombie world... in any case, we can bypass this worry altogether simply by considering a partial zombie world: one in which ... I am conscious with all the relevant concepts, but some people are zombies. (133)

If only the problem would dissipate so easily! But the presence of phenomenally-enriched minds in an otherwise-zombie-world tells us nothing about the *zombie's* epistemic situation here: does the zombie can have phenomenal concepts; if so, what would be their content? Can zombie-beliefs take phenomenal properties as their objects? If not, what should we make of the zombie's claim that she is conscious?²¹ When the zombie-dualist debates the zombie-eliminativist over the nature of consciousness, what is the subject matter of their dispute? When the zombie-dualist and zombie-a posteriori physicalist claim they can conceive of zombies, or when my zombie-twin writes her zombie-dissertation, what are those zombies talking and writing *about*? We, the phenomenally-blessed, have supposedly acquired our phenomenal knowledge and phenomenal concepts by being “acquainted” with phenomenal experience, by bearing an especially intimate epistemic relation to the content of our experience. How could zombie-“phenomenal” talk be about *consciousness itself* if they are wholly *unacquainted* with the phenomenon?

These questions have received little attention and the issue has remained largely unexamined. But this “zombie-phenomenal talk” deserves serious philosophical attention,

²¹ Per the description of the philosophical zombie above, we assume that they have propositional attitudes, though the objects of these propositional attitudes are explicitly non-phenomenal; whatever other restrictions there may be on the content of zombie propositional attitudes remains an open question.

not merely from dualists but physicalists as well. The a posteriori physicalist *agrees* that the antecedent of the conceivability-possibility thesis is true; zombies and inverteds are genuinely conceivable.²² And when the a posteriori physicalist conceives of zombies, she will conceive of them as uttering sounds that are phonetically identical to the sounds with which she expresses “phenomenal judgments” and “conceivability judgments” such as, “I am having a *red experience* now,” and, “I can conceive of someone physically identical to myself whose color-qualia are inverted relative to my own.”

What should the a posteriori physicalist make of the zombie’s “phenomenal” talk, of zombies deploying “phenomenal” concepts? To defend physicalism against the anti-materialist conceivability argument she must block the dualist’s move from the conceivability of zombies to their possibility, but, as an *a posteriori* physicalist, she must do so without denying its founding assumption, i.e., the conceivability of zombies, which each side takes to be an obvious truth and one of the features that distinguishes both from the “a priori physicalists”. The a posteriori physicalist does not have the option of claiming that zombies are merely *prima facie* conceivable; these creatures must be *ideally conceivable*, since any concession short of *ideal* conceivability is tantamount to denying the conceivability of zombies. And if zombies are ideally conceivable, we ought to be able to “fill in” all the details of the zombie-scenario, including the content, reference, and, generally speaking, the *meaning* of zombie-“phenomenal” concepts and beliefs.

²² According to these physicalists, the structure of our cognitive economy allows us to conceive of our phenomenally-deprived, physically-identical zombie-twins; that is, our phenomenal concepts are supposedly “isolated” from whatever non-phenomenal concepts we might have (physical, functional, structural, intentional concepts), and this “conceptual isolation” prohibits any a priori deduction of phenomenal facts from physical facts. We will discuss this in detail further along.

Though originally a seemingly miniscule detail of the overall zombie-story, to establish the genuine or ideal conceivability of zombies, the a posteriori physicalist must be able to give a consistent account of zombie-talk, “phenomenal” as well as non-phenomenal zombie-conversation. If she cannot find a way to accommodate zombie-“phenomenal” concepts in her overall metaphysical position, the a posteriori physicalist will need to reconsider her conceptual commitments; in doing so, her position will inevitably move closer to a kind of dualism or toward a priori physicalism.

The success of the property dualist’s conceivability argument does not depend on our *actual* ability to fully imagine our phenomenally-deprived, physically-identical twins, but its success does require that all details of the zombie-scenario can, *in principle*, be filled in without disrupting its conceptual coherence. Even so, the conceivability of zombies will be of no use to the property dualist’s argument if we find that the epistemic situation of these phenomenally-deprived creatures is no different from our own. Conversely, if the a posteriori physicalist cannot show that zombies *do* share our epistemic situation, she will be at odds to distinguish her metaphysical position from either property dualism or a priori physicalism. Questions about zombie-minutia such as “zombie-phenomenal concepts” have become relevant to both the property dualist’s defense of dualism and the tenability of a posteriori physicalism.

It is time to take zombies, and zombie-“phenomenal” concepts, seriously.

1. A Little Background

1.1 Qualia: Inverted and Otherwise

Imagine me and you, sitting in a park, watching a blue hummingbird dogfight with his yellow compatriot, both darting at a feeder full of their favorite food- bright red simple syrup. We are both admiring the bold and beautiful contrast of colors- scarlet red water, blue and yellow birds, all popping out against a green grassy background. We are both fully *phenomenally* conscious of this colorful scene, and for a moment you entertain the thought (as many of us did as children), “Does she see this scene the way I see it? Does the blue hummingbird look blue to her as it does to me? Perhaps when she sees the blue hummingbird it looks, to her, the way the *yellow* bird does to me. What is it like for her to see the birds at the feeder? Is her experience just like mine, or could it be just the opposite?”

There is nothing in the story above that would cause problems for physicalism. If there are differences between how you and I see the birds and the trees, it may be due to some physically-reducible or supervenient functional difference in our visual systems. My vision may be impaired due to some physical deterioration, or you may be jaundiced and your vision tinted slightly yellow. The dogfighting hummingbirds may not look *just the same* to me as they do to you because of physical differences between us that underlie functional differences in our color vision.

The *real* inverted qualia thought-experiment involves *idealized twins*, call them Brian

and Ryan, once again molecule-for-molecule and functionally identical to each other, both fully conscious, both receiving the same sensory input from the world, processing it physically identical ways, and producing the same output, whether that “output” is external behavior or changes in their brain states. The frequencies of the light-waves that hit Brian’s eyes is the same as those that reach Ryan’s, the images on their retinas are identical, and their resulting brains states are physically and functionally identical. And for each of them there is something that its like to see the rose on the bush; each has an experience with determinate qualitative content.

But these twins are Inverts- when Brian and Ryan look at the same rose in front of them, under their idealized identical conditions, the qualitative contents of their rose-experiences are *inverted relative to one another*. And as *idealized* twins, no difference between Bryan’s experience and Ryan’s can be explained by physical or functional differences between the two. The contemporary *loci classici* of the Inverted Qualia thought-experiment are works by Ned Block (1976, 1992). He writes,

It makes sense, or seems to make sense, to suppose that objects we both call green look to me the way objects we both call red look to you. It seems that we could be functionally equivalent even though the sensation fire hydrants evoke in you is qualitatively the same as the sensation grass evokes in me. (1992: 81)

Those who find this perfect “inverted qualia” story conceptually coherent claim that while the physical and functional facts about Brian and Ryan may be identical, it is still *conceivable* that some of the content of their conscious experiences could differ. *What it’s like* for Brian to experience the color of the red rose may not be identical to *what it’s like* for Ryan to see the same rose. Brian’s phenomenally conscious color experience of the rose may be identical to Ryan’s phenomenally conscious color experience of a blueberry. When Brian

has the conscious experience of seeing something paradigmatically red he has the experience that Ryan has when he sees something paradigmatically blue. Brian's and Ryan's phenomenal-red and phenomenal-blue experiences –their red and blue qualia- are “inverted”, though the twins are molecule-for-molecule physically and functionally identical.

Consider a different but related example. You are undergoing brain surgery; conscious, though locally anesthetized, you are conversing with your doctors while they probe different points in your brain. When a certain fold of your gray matter is prodded your speech begins to slur; another poke and you tell the doctors that you're having the experience of a lemony-taste.

There is little mystery to the connection between the manipulation of your gray-matter and your slurred speech. If the surgeons were to tell you that the function of that bit of the brain was to regulate the motor system responsible for speech production, and tampering with that area disrupted normal motor control and resulted in your slurred speech, you would probably feel that everything that *could* be explained about the brain/slurred speech connection had been explained. The whole process is a motor process, a physical process, and the brain is part of that physical chain of events. Such an explanation, were it given in sufficient detail, would seem satisfactory in the way that, borrowing from Kripke's examples, the explanations of heat as molecular motion and lightning as electric discharge also seem satisfactory. All the potential “why questions” seem to have been answered.

The surgeons might also give you a similarly functional explanation of the connection between the second brain-poke and your “lemony-taste” sensation, that is, a fully physical explanation of how a bit of brain stimulation could be responsible for your subjective experience of a “lemony” quale. But this second explanation might seem more mysterious.

There may be a similar causal-chain story, starting with the gray matter and ending in your experience of the lemony-taste; such a story seems to be as much of an answer as a physical explanation of the ‘brain-to-lemony-experience connection’ can provide. A causal-chain story may answer the “how” questions relevant to your “lemony-taste” experience; it may tell you *how* the brain-prodding caused your lemony experience. But unlike the surgeons’ explanation of your slurred speech, there seems to be a hole in the explanation of the physical brain’s connection to your experience of the lemony taste; a disappointing “gap” between brain process and subjective experience. There seem to be several legitimate questions left open: we can ask not only *how* the event occurred, but also *why* poking that part of the brain gave you the experience of lemony-taste rather than the experience of sugary-taste, or venison-taste, or no taste whatsoever. Why did poking your brain give you *any sensory experience at all*? Couldn’t the chain of physical events that started with prodding gray matter and ended in your report that you tasted lemons have skipped that subjective lemony-taste-experience “step”?

What are these properties of Brian’s and Ryan’s experiences that may be “inverted” relative to each other? And why should poking a lumpy gray mass in your brain result in a subjective experience of lemony-taste? What are *qualia*, such that they can invert, or disappear, though we may remain the same in *every other* respect, and, a further question, which “respects” qualify as “other”? What are these mental properties that, as Block says, “may well not be in the domain of psychology”?

Just what are the zombies supposedly missing?

1.1a Irreducible, Enigmatic, Problematic

The thought-experiments above are meant to coax our intuitions against *functionalism*, the theory that mental states, sensations included, are identified by their causes and effects, both overt and covert. This physicalist theory of mind seems to leave out an explanation of why all the whizzing and whirring of brain activity- the activity that creates intelligent *behavior*- should be accompanied by sensation, by qualia, by a rich inner life inside the mind. Qualia seem to defy functional explanation. Nearly every philosopher who writes on consciousness has a distinct account of the nature of qualia: according to some, qualia may be compatible with functionalism and physicalism, and according to others, qualia are of their own ontological kind. Questions arise as to what qualia are properties of, what *order* of property might they be, and, at an extreme end, whether “qualia” is a useful notion at all. Would philosophers of mind be better off if they discarded the term entirely?

Probably. But since qualia simply *are* what the zombies are missing we need a general understanding of this mired terrain. We could try to put the matter simply, and say, to borrow from Nagel (1974) that there is *something its like* to have conscious experiences-- to here a loon call on a lake, to taste slightly-too-hot hot chocolate, to see a flash of lightning, and qualia constitute the “*what it's like*”-ness of the experience. David Chalmers gives a florid description of our quarry:

The subject matter is perhaps best characterized as “the subjective quality of experience.” When we perceive, think, and act, there is a whirl of causation and information processing, but this processing does not usually go on in the dark. There is also an internal aspect; there is something it feels like to be a cognitive agent. This internal aspect is conscious experience. Conscious experiences range from vivid color sensations to the elusive experience of thoughts on the tip of one’s tongue;.. from the specificity of the taste of peppermint to the generality of one’s experience of selfhood. All these have a distinct experienced quality. All are prominent parts of the inner life of the mind. (1996: 4)

Qualia are sensations, or sensory experiences, or properties of sensory experiences.²³ They travel under many titles, but generally speaking, qualia constitute the phenomenal character of our conscious experience. Qualia are what we introspect when we focus on the subjective character of the experience of smelling a rose, or hearing a bird call. Michael Tye (2003) defines qualia as simply, “the introspectively accessible, phenomenal aspects of our mental life.” In this very neutral sense, “it is difficult to deny that there are qualia” (ibid: 1).²⁴

Beyond this very thin characterization, accounts of “qualia” quickly diverge. The widest division is between the physicalists, according to whom everything that exists is either straightforwardly physical or supervenes upon physical properties as, e.g., tartan *patterns* that supervene upon sets of (physical) colored threads, and the dualists, according to whom qualia are of a unique ontological kind that neither reduces to nor supervenes upon any *non-*phenomenal properties, i.e., physical, functional, or representational properties. And within these two divisions, uses of the term continue to differ.

²³ Qualia are not necessarily *limited* to sensation. Intentional states may have phenomenal content or qualitative character (suggested here by Chalmers’s mention of the “elusive experience of thoughts on the tip of one’s tongue”). This has been a subject of much interest, and, as Lycan writes, “the issue is just that of whether there is nonsensory “phenomenology,” and in particular whether “the terms *qualia* and *qualitative*” should be restricted to sensory states” (2008: 4). (Lycan attributes the first suggestion of phenomenal intentionality to Alvin Goldman (1993).) I will generally restrict the use of “qualia” and “phenomenal property” to sensory states, though considerations of “phenomenal intentionality” could play into the connection between phenomenal and cognitive mental states.

²⁴ The notion of a “quale” (the singular of “qualia”) has a rich legacy. C.I. Lewis used the term in 1929 to refer to qualities of sense data, but its history reaches much further back. Leibniz had a notion of qualia, something that could conceivably be missing from a “machine” functionally identical to a conscious subject (though he used the term “perception” to refer to it). As he wrote,

One is obliged to admit that *perception* and what depends upon it is *inexplicable on mechanical principles*, that is, by figures and motions. In imagining that there is a machine whose construction would enable it to think, to sense, and to have perception, one could conceive it enlarged while retaining the same proportions, so that one could enter into it, just like into a windmill. Supposing this, one should, when visiting within it, find only parts pushing one another, and *never anything by which to explain a perception*. Section 17 of the *Monadology* (1714)

William Lycan describes at least seven uses of “qualia” that appear in the literature, two of which are the most popular. The first kind of “quale” is what he calls a “Q-property”,

an introspectable (apparently) monadic qualitative property inhering in a mental state, such as: the color occupying such-and-such a region of your ordinary visual field right now... For example, if you experience a green after-image as a result of seeing a red flash bulb go off, the greenness of the after-image is the Q-property. (2008: 10)

Q-properties, as Lycan describes them here, are metaphysically neutral- they may be either physical or non-physical properties. One can endorse “Q-property”-qualia without taking on any particular ontological commitments. You and I are familiar with “after-images”- you can attend to them, describe their appearance, intentionally induce them, etc. These introspectable qualities of our conscious experience- these “qualia”- are uncontroversial; “*what it's like*” to experience an after-image is simply *having an experience with qualities such as these*.²⁵

“Qualia” in the second, stronger, sense, are the philosophically controversial qualia.

These are the potentially ontologically unique properties of conscious experience, the qualia

²⁵ Physicalists deny qualia in the upcoming, stronger, sense, but even those who claim that experience is “transparent” could endorse qualia in the first, milder sense- the “Q-properties”. G.E. Moore takes this position when he writes, “When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous” (1903: 450). We see *through* the experience, so to speak, to the object that is the source of our sensation. “The blue” in Moore’s example qualifies as a quale in the mild sense. On Tye’s (2000) account, qualia are *intentional objects*; they are simply *the way that external objects are presented in perception*.

You are immediately and directly aware of a wide host of qualities. You experience these qualities as qualities of the surfaces [of the objects in front of you]. You do not experience any of these qualities as qualities of your experience. There are no qualities of the experience that one is aware of; one is simply aware of the qualities of the objects seen. The experience of seeing is transparent. (45)

Gilbert Harman (1990) defends a similar position:

When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience... Look at a tree and try to turn your attention to the intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree. (39)

that do not logically supervene upon or reduce to physical properties. A “green quale” in this sense of “qualia” is not a quality of the green after-image *itself*, but a quality *of your experience of your after-image*, that is, *what it is like* for the subject *to experience* a particular quale;

[W]hat was it subjectively like for you to experience the greenness? Whatever it was like was a higher-order property of the greenness itself: Some philosophers, such as Block (1995) and Rey (1998), have used “quale” to mean this higher-order property, the “what it’s like”-ness that is a feature of a Q-property. (ibid: 10)

“Qualia”, in Block’s controversial characterization below, are neither physical, nor functional, nor representational properties. Block here divides mental content into two, non-exhaustive, types,

one of which is a matter of the way the experience represents the world, the other of which is a matter of “what it is like” (in Tom Nagel’s phrase) to have it. The former has been called intentional or representational content; the latter qualitative or sensational content. (1990: 3)

Block reserves the term “qualia” for the second type of mental content and suggests that “qualia may well not be in the domain of psychology,” (1990: 3) implicating neuro-psychology and any of other the “special”, physically reducible, sciences.

Here, Lycan and Block are careful to distinguish between the different uses of the term, but most appearances of “qualia” in the literature are ambiguous, leaving the reader uncertain about which of the different uses is being employed²⁶. One example of this ambiguous usage- one among many- appears in Smythies (2008),

‘Qualia’ may be defined as ‘what it’s like’ to have a particular experience. For example, it is held that the difference between a blue quale and a red quale is

²⁶ There are many more uses than these two described here; Lycan recognizes at least six. But for the sake of this discussion we can limit the “types” of qualia to these to general categories.

fundamentally intrinsic and *cannot be fully accounted for* only in terms of physical items such as differential wavelengths of light, or differential activation of neurons in the cortex. (191, emphasis mine)

The claim that qualia “*cannot fully be accounted for only in terms of physical items,*” might merely mean that we cannot give an explanation of qualia cannot in physical *terms* or that qualia cannot be *conceptually* reduced. Most physicalists (a posteriori physicalists) would concur. But the statement could equally be interpreted as the stronger claim that qualia *do not supervene upon* and *therefore* “cannot be fully accounted for” in terms of physical *items*, i.e., physical *properties*. Physicalists would, of course, dissent.

Disagreements over *how* zombies are to be conceived (over *what* they lack when they lack qualia) often arise because interlocutors use “qualia” in incommensurate or ambiguous ways—this might be an argumentative strategy, but more likely it is only “loose talk” which, at the end of the day, will not move the discussion of consciousness- or zombies- forward. This ambiguous usage of the term “qualia”, and the “wheel-spinning” it engenders, motivates some- notably Daniel Dennett- to claim that *there are no qualia at all*, and he (as well as Lycan) suggest that we dispense with the term entirely. In “Quining Qualia” Dennett writes,

My claim, then, is not just that the various technical or theoretical concepts of qualia are vague or equivocal, but that the source concept, the “pretheoretical” notion of which the former are presumed to be refinements, is so thoroughly confused that even if we undertook to salvage some “lowest common denominator” from the theoreticians’ proposals, any acceptable version would have to be so radically unlike the ill-formed notions that are commonly appealed to that it would be tactically obtuse—not to say Pickwickian—to cling to the term. Far better, tactically, to declare that there simply are no qualia at all. (1990: 520)

1.2 The Metaphysical Clubs

1.2a Property Dualism- Getting Metaphysical Mileage out of Epistemic Evidence

The heated qualia-debate concerns qualia in the “strong sense”; non-physical, non-functional, non-representational, higher-order properties of conscious experience. These qualia are the “intrinsic features of experiences which can vary without any variation in the intentional contents of the experiences” (Tye, 2003: 1). When property dualists argue that qualia exist and that zombies are metaphysically possible, they are arguing for the existence of qualia in the “strong” sense—this is the kind of qualia that we, as conscious human beings, supposedly have and our zombie-counterparts are supposed to lack.

Maund (2008) advocates for the existence of these “strong” qualia, which are both “introspectively accessible features of experiences...and... are intrinsic, non-intentional features” (271). His suggestion is not entirely unreasonable. If someone were to ask, “How does this tea taste to you?” you might respond that it tastes fruity, or smoky, or bland, etc. If you’re asked *what it’s like for you to experience* that fruitiness, you might try to answer by saying that it’s pleasant, or too tart. But there is a strong intuition that any answer of this sort will leave something out, that there is some ineffable feature of your conscious experience, a feature not explicable in functional, physical, representational or intentional terms.

According to the property dualists, these ineffable, seemingly *non-physically-explicable* features of your conscious experience are non-physical mental properties—“phenomenal properties” or “qualia”. They are ontologically distinct from your psychological or cognitive mental properties. They are non-intentional, non-representational, non-functional, and they compose the qualitative character of your experience. To have a

conscious experience with a red quale *just is* to have a red experience. Qualia only exist insofar as they are instantiated in a subject's experience, and if your experience has a red or green or lemony quale, you will know it. As Chalmers writes,

There is not even a conceptual possibility that a subject could have a red experience like this one without having any epistemic contact with it: to have the experience is to be related to it in this way (1996: 197).

“This *way*” in which you are epistemically related to your experience, the kind of epistemic contact that you have with your qualia, is more intimate than any other epistemic relation. It is “acquaintance”, and this epistemic relation holds *only* between you and the qualitative properties of your experience. Acquaintance brings you closer to the object of your knowledge than any other relation could possibly bring you.²⁷

The dualists claim that these qualia, and this kind of epistemic relation- acquaintance- are physically inexplicable. Acquaintance cannot possibly be a physical relation- a relation between two physical entities²⁸. And a physicalist conception of the mind does not have the resources to explain *why* our experience has a qualitative character, why qualia should exist along side the physical, functional, representational properties of our mental states.

1.2a1 We Can't Tell You Why... Returning to the “Gap”

As the apparent conceivability of zombies shows, we can conceive of the physical processes going on in the absence of sensation, in the absence of an “inner life”. Why doesn't

²⁷ One peculiarity of “acquaintance” in this sense is that it implies that your knowledge about the qualitative character of your experience is more intimate, and more certain, than your knowledge of your *beliefs about* your qualia. This seemingly paradoxical feature of this notion of “acquaintance” is the subject of other work in progress and will not be explicitly discussed here.

²⁸ This is acquaintance as the dualist defines it; there are other accounts of our epistemic access to our conscious states that also go by the name “acquaintance” and are neutral between metaphysical theories of mind.

all this brain activity, this information processing, happen “in the dark”? If the mental is physical, if the mind is the brain in action, if they are either identical or the former supervenes upon the latter, why do the mind and brain seem to be so radically different, and why do they continue to seem radically different despite whatever new discoveries are made about connections between particular brain processes and particular mental states?

The explanatory gap is opened by the conceivability of the mind as separate from the brain, and this gap between physical and phenomenal facts, between qualia and brain states, has been with us as long as anyone has wondered whether the mind could or could not be a physical entity.²⁹ The gap separates Cartesian immaterial and material substances, a separation born of Descartes’ ability to conceive of the immaterial substance apart from material substance, as it seemed possible for either to exist entirely without the other.³⁰

Philosophers have widely rejected Cartesian substance dualism, as it is at least as rife with problems as any other theory of mind, the most obvious being its mind-body interaction problem. How could substances entirely distinct in *kind* interact or have causal influence on each other? The senses, themselves wholly physical mechanisms, must gather information about the world to be transmitted to the immaterial mind where conscious experience occurs. Descartes’ suggestion that the pineal gland could be the locus of this interaction appeared *ad*

²⁹ Contemporary philosophers phrase the problem in terms of explanation: the kind of explanations that a physicalist account of the relationship between mind and body supposedly lack are “transparent explanations”, explanations that can,

make transparent why relevant high-level truths obtain, given that low-level truths obtain. If it is conceivable that the low-level truths obtain without the high-level obtaining, the explanation will not be transparent in the relevant way. Instead, one will need to appeal to substantive further principles to bridge the divide between the low-level and high-level domain. It is just this sort of transparent explanation that is absent in the original explanatory gap. (Chalmers, 2010: 395-360)

³⁰ Meditation VI

hoc, and the eventual acceptance of the causal closure of the physical world allowed scientifically minded philosophers, Empiricists, to dismiss substance dualism as a viable theory of mind.

But the conception of the mind as distinct from the body remains, and “the principle of the causal closure of the physical world” is merely a bridge between the mind and brain- a tightrope across the explanatory gap. “Causal closure” does not *close* the gap, and though to some philosophers it has seemed to justify widely *ignoring* the gap, one can endorse physicalism while taking the explanatory gap seriously, viewing it as a genuine though potentially resolvable philosophical problem. Such is the view of Joseph Levine, the physicalist to whom we attribute the phrase. Locke, the empiricist, wondered how the body could be the mind, and Leibniz pondered his windmill. But they refrained from returning to Descartes’s failed attempt at dividing the world into types of substance, the mental and the physical. Their inquiries into the mind focused on questions about memory and perception rather than the “ultimate” question of the connection between the mind and the body. But the explanatory gap is *the issue* in the mind/body problem;³¹ the gap itself is epistemic, a problem about how we can know, or whether we can explain, facts about our subjective experience by appealing to physical facts alone.

1.2a2 Mind-Splitting- Phenomenal and Cognitive Mental Properties

Property dualists hold that the apparent epistemic gap between mental facts and physical facts cannot be accounted for by appeal to physical and functional properties alone,

³¹ At least, this is the view of Levine (2007); many philosophers of mind would likely agree.

and that some sort of metaphysical dualism is responsible for its presence. But unlike Cartesian substance dualism, this ontological divide between conscious experience and the physical is not limited to a separation between mind and world; it creates an ontological rift within the mind itself. According to the property dualists, our minds consist of two ontologically distinct and conceptually independent kinds of properties: psychological or cognitive mental properties (which are reducible to or metaphysically supervene upon ordinary physical and functional properties) and “phenomenal” mental properties, or “qualia”.

Psychological properties serve as the “causal or explanatory basis for behavior... what it means for a state to be psychological is for it to play an appropriate causal role” regarding overt behavior or covert behavior such as reasoning or introspecting (Chalmers 1996: 11-12).³² These mental properties include common mental functions such as attending, voluntarily controlling one’s actions, wakefulness, knowing, remembering, reporting, identifying, etc. According to property dualism, these psychological or cognitive properties do not constitute conscious experience in the relevant sense. They are mere functional states; as functional states logically supervene upon our physical states, our zombie-twins will possess these functional psychological states just as we do.

Phenomenal properties, sensations, raw feels, *qualia*, neither reduce to nor logically supervene upon physical or functional properties (ibid: 124). As Chalmers explains,

conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those

³² It is controversial whether some of these “psychological” mental functions require a phenomenal component, for instance, whether zombies will be able to “introspect” in the full sense of the term. We will leave this issue aside for now but it may surface in later discussion.

properties. Consciousness is a *feature* of the world over and above the physical features of the world... there are properties of individuals in this world—the phenomenal properties—that are ontologically independent of physical properties. (1996: 125)

Naturalized property dualists such as Chalmers assume there is a nomological connection- something like *strong emergence*- between conscious experience and certain physical states such that, in our world, conscious experience does not occur without some connected psychological feature. But because these are ontologically independent properties, it is logically possible for phenomenal and psychological properties to vary independently of one another. According to property dualists, it is neither metaphysically or conceptually necessary that any particular psychological or cognitive property accompany the properties of a conscious experience. The *phenomenal* properties of a conscious experience are its defining feature; they are the “intrinsic” properties of that experience, the properties that make a conscious experience *the experience that it is*. “What it *means* for a state to be phenomenal is for it to feel a certain way” (Chalmers, 1996: 12, emphasis original).

A conscious experience will have certain “background” properties; these properties are extrinsic but related to the experience itself. Psychological properties fall within this category, as do the physical brain states that are nomologically connected to the conscious experience, and the relations these phenomenal mental states bear to the environment, i.e., occurring at a certain time, in a certain place, to a certain subject, etc. Kripke (1972) argues for a conceptual distinction between these background properties and, for example, the phenomenal quality of *pain*. If the property dualist has the proper conception of mental properties, then the phenomenal property ‘painfulness’ can be cleanly separated from its commonly associated “accidental” properties. Behaviors such as yelping, recoiling from the

supposed stimulus of a pain sensation, the disposition to favor the part of the body in which the pain seems to be located, are “accidental” properties of the phenomenological pain-experience itself, “background” properties.

As Kripke writes, “Pain... is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality” (ibid: 152). The functional, behavioral, physical, and psychological properties that are normally co-instantiated with the phenomenal property ‘painfulness’, including your *belief* that you are in pain, might all be present in a particular mental event FP, but, in principle, the phenomenal property ‘painfulness’ could itself be absent from the mental event FP, in which case the FP will not be a pain-experience, despite your pain-type reactions and your protests to the contrary. “Painfulness” may be the most contested case of a purely phenomenal property, but if the notion of a zombie is coherent there will be a coherent notion of “zombie-pain”, the zombie-state that corresponds with an “actual-pain” state, in which all these background or accidental psychological features of pain are present in the absence of *actual* ‘phenomenal-painfulness’.³³

1.2b The Physicalists

Recall the simplified version of the conceivability argument introduced earlier:

1. Zombies are conceivable
2. Whatever is conceivable, is possible
3. Zombies are possible
4. Physicalism is false

³³ See Lewis (1980) for a compelling discussion of the complexities of the functional and phenomenal properties of painfulness, including “mad pain” and “Martian pain”.

Because philosophical zombies are defined as *void of conscious experience*, different interpretations of the meaning of *consciousness* and *conscious experience* will generate different conclusions about the conceivability of zombies. These differences not only separate physicalists from dualists but create division within the ranks of the physicalists themselves. Both dualists and a majority of physicalists (“a posteriori physicalists”) interpret “having a conscious experience” as something like “being in a state that has a certain phenomenal feel” or “there being *something it’s like* to be in this particular state.”³⁴ There is a smaller set of physicalists, “a priori physicalists”, for whom the role of “seeming” is essential to the notion of *conscious experience*; for these a priori physicalists, for a subject to *have a conscious experience* means that subject is in a state that *seems to the subject* to have a certain character to it (you might say, the state *seems to the subject* to have a certain “quality” or “feel”).³⁵

1.2b1 “A Priori” Physicalism: A Straightforward Rejection of “Common” Sense

If “seeming” plays an essential role in the concept conscious experience itself, the definition of a philosophical zombie will be incompatible with the concept of conscious experience, since it will certainly *seem, to the zombie*, that she is conscious and has conscious experiences. In the description of zombie-worlds, zombies claim to be conscious, they write

³⁴ We will note here that these “meanings” are only the primary intensions of conscious experience on a two-dimensional semantic framework. A posteriori physicalists and property dualists do not agree on the entire content of the concept conscious experience- they disagree on its secondary intension, what the term picks out in the actual world (a physical or phenomenal property). We will discuss this in more depth shortly.

³⁵ According to a smaller set of physicalists, perhaps eliminativists, “conscious experience” may be so highly ambiguous that it means nothing at all. Though many interpret Dennett’s view on “consciousness” in this way, it would be a mistake to do so. As we will see in Part two, Dennett does hold that the ambiguous usage of the term “qualia” has rendered “*qualia*” meaningless, but “qualia” should not be equated with either “consciousness” or “conscious experience”- one term may be rejected without rejecting the others.

about consciousness, complain when they feel *that pain* in their lower back again, etc. All parties to the zombie-debate agree that there is a philosophically significant sense in which it *seems to the zombie* that she has conscious experience when in fact she does not. But if it *seems* to the zombie that she is having a conscious experience, then- according to this interpretation of “consciousness”- the zombie *is* having a conscious experience and thereby conscious. Thus the notion of a philosophical zombie contradicts the a priori physicalists’ concept of consciousness; if all that is required for consciousness is for it to *seem* to the subject that she is conscious, then zombies are conceptually incoherent, i.e., inconceivable.³⁶

“A priori” physicalists are physicalists who deny the primary assumption on which the conceivability argument is based: its first premise, that zombies are genuinely conceivable. The apparent conceivability of zombies may stem from a conceptual confusion about the nature of consciousness, or latent disagreements over what it means for a scenario to be conceivable, or from a “lack of imagination”; in some way or another, the *description* of a “philosophical zombie”- itself so easily intelligible- is obscuring the actual complexity of the zombie-scenario, and its complexity in turn hides the zombie’s conceptual incoherence. Zombie-stories induce powerful conceptual hallucinations in which what you seem to be conceiving isn’t really what you conceive. According to some physicalists, this “condition” might be cured by the discovery of a yet-undiscovered but in-principle-possible a priori connection between phenomenal facts, i.e., facts about conscious experience and physical facts, hence the name, “*a priori*” physicalism. Zombies, so these physicalists claim,

³⁶ Accounts according to which all that is required for a subject to be conscious is for it to seem to a subject that she is conscious are (generally) categorized as “anti-realist” accounts of conscious experience.

are not *actually* conceivable. The zombie is not a conceptually coherent notion, though descriptions of zombies may well be intelligible.

1.2b2 “A Posteriori” Physicalism: The “Common” Sense- Grant Conceivability, Deny Possibility

Dualists and the majority of physicalists take the meaning of “consciousness” to be a something like “having phenomenal feel”. This group of physicalists, “a posteriori physicalists”, endorse premise 1 of the conceivability argument: zombies or zombie-scenarios are obviously and unproblematically conceptually coherent. According to a posteriori physicalists, there is no conceptual contradiction within the zombie-scenario, the scenario in which it *seems* to a subject that she is having a conscious experience with a certain phenomenal content although she is having no conscious experience whatsoever. If “having a conscious experience” means “being in a state with a certain phenomenal feel”, and *seeming* to be in a state with a phenomenal feel is not conceptually *sufficient* for having a conscious experience (veridical or otherwise), then a subject may *believe* that she is in a state that has a phenomenal feel, but holding this belief will be conceptually consistent with the subject not actually being conscious. The subject may hold a false belief about her mental state,³⁷ but the scenario is itself is conceptually consistent.³⁸

But the a posteriori physicalist insists that zombies are *merely* conceivable, and premise 2 of the conceivability argument is their standard target. Facts about possibility and

³⁷ Or her belief may be meaningless, or she may hold no belief whatsoever, or she may be in some other sort of state; we will discuss all such options in due course- the important point here is simply that the scenario is conceptually coherent, not the status of the subject’s “belief”.

³⁸ Views on which zombies are conceivable- conceptually coherent- are (generally speaking) “realist” accounts of consciousness, while those that deny the conceivability of zombies are (generally speaking) considered to be “anti-realist” accounts of consciousness.

facts about conceivability are distinct though related matters, and a posteriori physicalists reject the claim that the conceivability of scenario S straightforwardly *entails* that S is possible: conceivability is a *guide* to possibility, but a defeasible one.

1.3 Can the Metaphysically Impossible be Genuinely Conceivable?

1.3a A Posteriori Physicalism and Kripke's "New" Necessity: A Plurality of Possible Worlds

Premise 2, "whatever is (ideally positively) conceivable, is possible," is the vital premise in the conceivability argument; its success depends on the inconceivability of metaphysical impossibilities. Some a posteriori physicalists believe premise 2 is undermined by Kripke's notion of *a posteriori necessity*. As Chalmers remarks,

[there is a] familiar class of purported counterexamples [that] arises from Kripke's analysis of the necessary a posteriori. It is often said that sentences such as 'water is not H₂O' provide counterexamples to the claim that conceivability entails possibility: it is conceivable that water is not H₂O but it is not metaphysically possible. (2010: 145)

According to a posteriori physicalists, we can conceive of zombies not because they are possible but because the *structure* of our cognitive economy *isolates* our phenomenal concepts (the concept of *what it's like to see red*, for example) from our physical, functional, cognitive, and otherwise *non-phenomenal* concepts (the physical concept *brain event b*, or the cognitive concepts *focusing attention* and *identifying a percept*). Our ability to conceive of zombies reflects only the nature of our concepts; just as our ability to conceive of water as distinct from H₂O does not entail the metaphysical possibility that water is not H₂O, our ability to conceive of zombies says nothing about the metaphysical possibility of zombies (or

the ontology of the actual world). According to these physicalists, the conceivability of the zombie-scenario does not itself show that physicalism is false in our world.

The widely accepted view among physicalists is that physicalism is contingently true of our world; the mental facts of our world supervene upon the physical facts, but physicalism is not the only conceptually coherent theory of the relationship between mind and world and it is conceivable that physicalism might have been false of our world. A smaller set of physicalists (and a large portion of anti-physicalists) treat physicalism as a necessary or conceptual truth (if it is true at all).³⁹ When the thesis is posed as a claim about *entailment*, that the totality of the physical facts of our world entail all the facts in our world, physicalism does seem to be a claim about necessity.⁴⁰

I will not try to settle this matter here (i.e., whether physicalism should be interpreted as a claim about the necessary relation between physical and mental states or a claim about what is in fact true of our world but only *contingently* so). Whether the psycho-physical conditional, $P \rightarrow Q$, is intended to be taken as necessary or a contingent truth of our world, the fact that certain mental states supervene (or are identical to) certain physical (brain) states certainly *appears* to be contingent. In either case, those who interpret physicalism as a necessary truth or contingently true can and often do appeal to Kripke's (1980) introduction of "a posteriori necessity" to defend their version of physicalism.

To borrow Frege's familiar example: the terms "Morning Star" and "Evening Star" both refer to the planet Venus, but both terms pick out their referent through different "modes of presentation" or different descriptions (or we can say the two terms have different Fregean

³⁹ See Davidson (1970).

⁴⁰ Either this, or all the physical truths of our world, if there is a difference.

“senses”). “Morning Star” picks out its referent by something like the description *the last heavenly body we see in the morning*, while “Evening Star” picks out its referent as *the first heavenly body we see in the evening*. The difference between the way these two terms pick out their shared referent indicates (or is) a difference in their meaning, a difference between the content of the concepts *Morning Star* and *Evening Star*.

As with every instance of identity, the Morning Star is the Evening Star *necessarily*. Still, the identity of the Morning Star with the Evening Star may *appear* contingent because this identity is not an a priori truth, that is, the fact that the Morning Star is the Evening Star is not true in virtue of the *content of the constituent concepts- Morning Star and Evening Star- alone*. As an empirical rather than conceptual truth it will seem as though conditions could have turned out differently- it *seems* as though the Morning Star might have turned out to be some heavenly body other than Venus. But an identity discovered a posteriori is *identity* nonetheless; it is necessary, true in every possible world.

A posteriori necessary truths will appear contingent because we must learn something about the world, we must gain empirical knowledge, in order to see their truth.⁴¹ A person who does not know that the Morning Star and Evening Star are Venus, who does not know that ‘the last heavenly body we see in the morning’ is the very same *thing* as ‘the first heavenly body we see in the evening’, can conceive of the Morning Star and the Evening Star being distinct objects. For all this person knows, *they may well be* two separate entities.

⁴¹ According to Papineau (2009) these “truths” do not seem contingent; they simply seem *false*. “I think that the so-called ‘explanatory gap’ is simply a manifestation of an intuitive conviction that dualism is true. It’s not that mind-brain identities are hard to explain—they are simply hard to believe. When we consider a putative identity like pain = C-fibres firing, our intuitive reaction is simply that this claim must be false”

Whether or not one is actually aware that these “stars” are the planet Venus, there is still a sense in which it is conceivable for the Morning Star to be distinct from the Evening Star. Keeping to the “possible worlds” model of modality, we can express the divergence of conceivability from possibility by postulating two independent sets of worlds:⁴² the first containing only “*conceptually possible* worlds”, i.e., worlds that are merely conceivable, while the second set contains “*metaphysically possible* worlds”, the kind “worlds” with which the a posteriori physicalist is actually concerned. Because the identity relation is *metaphysically* necessary, the Morning Star will be identical with the Evening Star in every metaphysically possible world. But because their identity is *not a priori* it is not a conceptually necessary truth, and as such there will be conceptually possible worlds, or “conceivable scenarios”, or “*epistemically possible* worlds” in which the Morning Star and the Evening Star are distinct entities.⁴³

Appealing to two independent sets of possible worlds allows a posteriori physicalists to reject anti-materialist “a priori” arguments against physicalism, that the identity between physical and phenomenal states is not knowable a priori, and if there is no a priori deduction of phenomenal facts from physical facts physicalism must be false. Most physicalists are “a posteriori physicalists” in this sense: they claim that we can conceive of brain events occurring in the absence of phenomenal events even though the brain event and phenomenal event are identical, and we can conceive of zombie-scenarios even though they are impossible, for the same reason that we can conceive of the Morning Star and Evening Star

⁴² Conceivable worlds need not be metaphysically possible, nor must metaphysically possible worlds be *conceivable*, as we will discuss later when we focus on the connection between conceivability and conceiver.

⁴³ Phrased another way, there will be conceivable worlds in which the referent of “Morning Star” is distinct from the referent of “Evening Star”.

being distinct entities though their identity is metaphysically necessary. Just as the concepts *Morning Star* and *Evening Star* have different content but both refer to the planet Venus, the concepts we use to pick out our conscious experiences, such as the concept ‘*what it's like to see red*’ and the concepts we use to refer to brain or cognitive events e.g., ‘*identifying the red percept*’ have different content but actually refer to the same *properties*. The identity of the conscious experience ‘seeing-red’ with the brain or cognitive event ‘identifying-the-red-percept’ or ‘brain-state-b’ may be an empirical fact, known only a posteriori, and be metaphysically *necessary* nonetheless.⁴⁴

1.3b Defending CP: The Modal Minimalist

We might say that the a posteriori physicalist’s modal universe contains two types of worlds- the “merely conceivable” and the “genuinely possible”. The property dualist cannot accept this distinction between two sets of worlds: he must maintain his claim that conceivability entails possibility, which means showing how every conceivable world can be a metaphysically possible world. If this were not the case, the claim that zombies are conceivable would carry no metaphysical weight.

According to Chalmers, there is a perfectly fine sense in which an a posteriori necessary identity such as ‘water is not H₂O’ is conceivable that still fits with his claim that conceivability implies metaphysical possibility. Appealing to a two-dimensional analysis of the content of concepts, Chalmers claims that when we evaluate a scenario S for

⁴⁴ Throughout most of this writing, “metaphysically identical with” and “metaphysically supervening upon” can be used interchangeably. Though the difference between the identity relation and supervenience relation is profound, the claims I am making here apply to both more or less equally well. Where and when the difference is pertinent I will make the appropriate distinctions between the two.

conceivability, we must do so according to either S's primary or secondary intensions. In the a posteriori physicalists' "Kripke cases", the sense in which some scenario might be *metaphysically impossible* but nevertheless *conceivable* can be explained (or explained away) by specifying the intension according to which we are evaluating S for conceivability. If the world represented by scenario S is metaphysically impossible although S seems conceivable, it will be conceivable only when evaluated according to one intension and inconceivable when evaluated according to the other- the world will not be both metaphysically impossible *and* the scenario conceivable when S is evaluated on only *one* consistent intension.

Consider the expression, 'water is not H₂O'. 'Water is H₂O' is necessary, though the identity of water with H₂O is known only a posteriori. 'Water is not H₂O' is false in every possible world- it is metaphysically impossible for water not to be H₂O - but the meaning of 'water is H₂O' can be analyzed according to either its primary and secondary intension. When 'water is not H₂O' is evaluated according to an its primary intension- evaluated for *primary* conceivability- 'water is not H₂O' will be conceivable.

If an expression's primary intension is a function from possible worlds *considered as actual* (or "centered worlds") to extensions, and we reasonably assume that the primary intension of 'water' is "watery stuff", then 'water' (according to its primary intension) will pick out *whatever stuff around here* fits the description "watery stuff".⁴⁵ The *meaning* of 'water', according to its *primary* intension, is the same (i.e., "*watery stuff*") in every possible world. If "here" is Twin Earth, the *watery stuff* around here will be XYZ (not H₂O), and since

⁴⁵ As intensions are functions rather than descriptions, describing intensions in this way does not commit two-dimensional semantic to descriptivism. Intensions could be (and probably are) fixed by some appropriate causal connection, but talking in terms of *descriptions* simplifies this superficial exposition a bit.

the primary intension of ‘water’ picks out *whatever stuff around here fits the description “watery stuff,”* the primary intension of ‘water’ picks out XYZ on Twin Earth (when Twin Earth is considered as actual).

When we treat Twin Earth (or any non- H₂O world) as actual, ‘water is not H₂O’ will be true according to its primary intension, so we can say that ‘water is not H₂O’ is both “primarily conceivable” and “primarily possible”- conceivable and possible when we evaluate the sentence according to its primary intension.

If we think of the primary intension of a term as the concept’s “reference fixer”, we can think of the *secondary* intension of a term as picking out *whatever the primary intension picks out in the actual world*- it is a function from possible worlds considered as *counterfactual* to extensions. Since the primary intension of ‘water’ is ‘watery stuff’, and, in the (*actual*) actual world, only H₂O fits the description ‘watery stuff’, then ‘water’, according to its secondary intension, means H₂O.⁴⁶ Given that the actual world is an H₂O world, the secondary intension of ‘water’ is H₂O *in every possible world* (considered as counterfactual). So, according to its secondary intension, ‘water is not H₂O’ means ‘H₂O is not H₂O’; since it is both metaphysically impossible and inconceivable for something (H₂O) not to be *itself*, ‘water is not H₂O’ is impossible and inconceivable according to its secondary intension.

The sense of “conceivability” in which “water is not H₂O” is conceivable- the sense of conceivability doing work in the a posteriori physicalist’s counterexamples to the CP

⁴⁶ If one thinks there is a real distinction between- on the one hand- the relationship of the primary and secondary intensions for “natural kind terms” and- on the other- the relationship of the primary and secondary intensions of ordinary “semantically unstable” terms, one will likely treat ‘water’ as a rigid designator, such that its primary intension is fixed in *the actual* world (our H₂O world) and “rigidly designates” H₂O in every possible world. I mention this here only to note that my description of conceivability and possibility according to primary and secondary intensions here is intended to be lean enough to accommodate any particular interpretation of two-dimensional semantics.

principle (Premise 2 of the conceivability argument) is *primary conceivability* (since the primary intension of ‘water’ picks out (something like) “the watery stuff” in all possible worlds). But, as Chalmers claims, when ‘water is not H₂O’ is evaluated according to its primary intension, ‘water is not H₂O’ is also *primarily possible* (that is, there are metaphysically possible worlds in which “the watery stuff” is something other than H₂O). The only sense in which ‘water is not H₂O’ is impossible (false in every possible world) is according to its secondary intension, where ‘water’ means ‘H₂O’ (in possible worlds considered as counterfactual). But ‘water is not H₂O’ is also *secondarily inconceivable*-metaphysically impossible worlds are only conceivable when we equivocate between primary and secondary conceivability and possibility.

According to Chalmers, all Kripkian “necessary a posteriori” cases- cases in which scenarios describing metaphysically impossible worlds are in some sense conceivable- can be explained away by specifying the intension according to which the scenario is being evaluated. Considered in this way, Kripke’s introduction of a posteriori necessity does not *rule out* any genuinely possible worlds; rather, the apparently-conceivable-but-metaphysically-impossible worlds *are genuinely possible worlds*, but possible worlds described in inappropriate terms:

any conceivable situation in which it seems that water is not H₂O (a Twin Earth world, say) should better be described as a conceivable situation in which water is still H₂O but in which there is watery stuff around that is not H₂O. (Chalmers, 2010: 145)

If the only way to justify the claim that scenario S can be both conceivable and impossible is for one’s modal judgment to equivocate between the primary and secondary

intensions of S, then the physicalist's purported counterexamples, the "Kripke cases," fail as objections to the Conceivability/Possibility principle.

If the Conceivability/Possibility principle holds, and if the zombie-scenario is *ideally positively conceivable*, it follows that zombies are possible and physicalism is false. The dualist believes he can dodge the physicalist's "Kripkean" counterexamples- cases of conceivability without possibility- by distinguishing between *prima facie* and *ideal* conceivability and the primary and secondary intensions of the Kripkean scenarios. All sides agree that it is *prima facie* conceivable that water is not H₂O (or conceivable according to the primary intension of 'water'), but it is not ideally conceivable that water (itself) is not H₂O (it is conceivable according to the primary but not the (fixed) secondary intension of 'water'). Since ideal conceivability is the only variety that the dualist claims implies metaphysical possibility, the *prima facie* conceivability of the (metaphysically impossible) scenario "water is not H₂O" is not a counterexample to CP.

The Kripkean "a posteriori necessity" objection may not stop the property dualist's move from conceivability to possibility, but there is another well-known a posteriori physicalist objection to CP aside from the Kripke-cases discussed above. One is an argument from the isolation of phenomenal concepts from physical, functional, and otherwise non-phenomenal concepts, the "phenomenal concept strategy", which we will discuss in due course, after we have said a little more about the notion of conceivability.

2. Oh, The Modality!

2.1 The “Varieties” of Zombie-Conceivability, and Previous Unsuccessful Attempts to Undermine Them

Premise 2 of the conceivability argument- the controversial premise- is an instance of the (controversial) Conceivability/Possibility principle, the claim that whatever is conceivable, is possible. Spelled out in these simple terms, the Conceivability/Possibility principle seems clearly false; there are myriad instances of conceivable scenarios that are metaphysically impossible. For the most part, conceivability is a good but fallible guide to possibility. Why should conceivability be a *perfect* guide to possibility in the zombie case?

Chalmers’s defense of the Conceivability/Possibility principle (and the conceivability argument) involves dividing the notion of conceivability into several “varieties”: positing these distinctions within the concept of conceivability itself, Chalmers reformulates the CP in a way that avoids the most obvious objections. Rather than the overly simplistic (and clearly false) claim that could be read as “whatever *satisfies any criteria for conceivability* is possible,” the CP principle is the more refined claim that whatever is *ideally* conceivable is possible, or, more precisely,

CP: Ideal primary positive conceivability entails primary possibility

The “ideal” variety of conceivability above is contrasted with “prima facie” conceivability; the “positive” variety with “negative” conceivability, and “primary

conceivability with “secondary”.⁴⁷ The distinction between *prima facie* and ideal conceivability is intended to separate those scenarios that merely *seem* conceivable to a subject from those which are *actually* conceivable. The notion of *prima facie conceivability* is fairly straightforward: S is *prima facie* conceivable for a subject when S is conceivable for that subject on first appearances (Chalmers, 2010: 143).

The dimension of “positive” conceivability and “negative” conceivability is concerned with *creation* and *elimination*; a scenario is positively conceivable when it can be “created in (modal) imagination”, while “negative notions of conceivability hold that S is conceivable when S is not *ruled out*” (ibid: 143). A scenario is *prima facie negatively* conceivable when initial consideration of the scenario reveals no obvious conceptual contraction within that scenario or hypothesis. A scenario is *ideally negatively* conceivable only if “the hypothesis expressed by S cannot be ruled out a priori even on ideal rational reflection,” (2010: 143) and will fail to be ideally negatively conceivable when S is found to be *prima facie* conceivable but its *prima facie* conceivability is “undermined by further reflection showing that the tests that are criterial for conceivability are not in fact passed” (ibid: 144).

The notion of forming a “positive conception of a situation,” however, is somewhat more complicated, since it involves appealing to one’s *imagination*, but not “imagination” in any ordinary sense; S is positively conceivable when one can coherently modally imagine a

⁴⁷ On a two-dimensional semantic framework, a hypothesis (scenario, expression, sentence, description, or proposition) will have both a primary and secondary intension, and the conceivability of that hypothesis or scenario may depend on whether it is evaluated according to its primary or secondary intension. Though crucial to understanding how the CP stands up to the challenge of “Kripkean” cases of a posteriori necessity, primary and secondary conceivability are not the main focus of most arguments against the CP, and we can leave the discussion of this dimension until chapter two where we will examine the CP specifically in light of these Kripke cases.

situation that verifies S, that is, “when one can coherently imagine a situation in which S is the case,” (2010: 144) which is “to in some sense imagine a specific configuration of objects and properties,” (ibid: 145) and “fill in arbitrary details in the imagined situation such that no contradiction reveals itself” (2010: 145).

As the word suggests, imagination is often imagistic, or in some way sensorial, and of course there are genuinely conceivable scenarios that cannot be “imagined” in these ways. Descartes’ chiliagon, for example: in visual imagination, a polygon with 1000 sides is indistinguishable from a circle, though the two are clearly distinct and both are conceptually coherent. “Modal imagination” cannot be perceptual or sensory imagination, nor imagining from the first-personal subjective perspective. To “modally imagine a scenario” is not to imagine viewing some scene, but rather to mentally “arrange” a set of “objects and properties” in a certain way that would either verify or fail to verify a proposition that represents a (potentially) possible world.

Insofar as we can conceive of a difference between zombies and conscious human beings, the scenario cannot be positively conceived by comparing *perceptions* of the two creatures in imagination; a zombie world is identical to a human world *in every observable respect*. Neither we nor the zombies could observe that zombies lack phenomenal states. Zombies themselves cannot observe that zombies are in any way different from humans. Zombies, by definition, do not observe any difference between themselves and their human twins- one cannot observe a subjective state that is simply absent (thus the zombie cannot observe a *difference*).

Zombie-minds will be *prima facie positively* conceivable when a subject believes that all the details of the zombie-scenario can be filled in without revealing a contradiction, or

when a subject believes she can “imagine a situation with certain important features specified, notes that a situation of this kind appears to verify S, and judges that the remaining details are not crucial (Chalmers, 2002).

2.1a Were it only so Easy: “Modal Imagination” and the Unfortunate Irrelevance of the First-Person-Subjective Unimaginability of Zombies to the Tenability of the Conceivability Argument

Of course, mere *prima facie* positive conceivability does not suffice for possibility. “For the thought-experiment to yield the intended conclusion, this *prima facie* judgment must be correct, so that S is ideally positively conceivable” (2002: 8). A subject may make the *prima facie* judgment that she can fill in the details of scenario S though the world S describes is actually conceptually incoherent (as in the examples of “inconceivable objects” discussed below). Most objections to the conceivability of the zombie-scenario argue against its *positive* conceivability, attempting to show that zombies are only *prima facie* positively conceivable, that subjects who believe they are conceiving of a zombie are actually making a mistake of one sort or another in their imagining- possibly by “filling in” the scenario with *improper* details, or failing to notice “holes” in their imagined creature.⁴⁸

Popular attempts to explain away the zombie’s apparent conceivability appeal to possible psychological facts about our faculties of imagination: when we take ourselves to be

⁴⁸ For compelling arguments of this sort, see Kirk (2008), Dennett on zombies (1991, 2005) and on the knowledge argument (1991), as well as Marcus (2004) and chapter four, here.

imagining a philosophical zombie, we imagine (or refrain from imagining) different aspects of the zombie-scenario.⁴⁹ To quote Nagel's account of imagination, at some length,

We may imagine something by representing it to ourselves either perceptually [or] sympathetically... To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived it. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself... When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state; that is, we put ourselves in a state that resembles it mentally. At the same time, we attempt to perceptually imagine the non-occurrence of the associated physical state, by putting ourselves into a state unconnected with the first: one resembling that which we would be in if we perceived the non-occurrence of the physical state. *Where the imagination of mental features is sympathetic, it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination.* (1979: 175, fn. 11, emphasis mine)

It is no mean feat to show that the creature we ordinarily imagine when we believe we are conceiving of a zombie is not a philosophical zombie *proper* (but rather some sort of ersatz creature irrelevant to the truth or falsity of physicalism), or to give a compelling explanation of our apparent ability to imagine philosophical zombies via an account of what might be facts about our actual faculties of imagination. And I believe that any (or all) of these accounts might be the proper diagnoses of our cognitive mistakes. But the problem with these and other prior attempts to identify the ways in which we might fail to properly positively conceive of a zombie is that they are *not objections to Chalmers's form of the conceivability argument*.

⁴⁹ See Hill (1997), Hill and McLaughlin (1999), Block and Stalnaker (1999), and Nagel (1974, fn11), among others. For an overview of this type of response to the conceivability argument, see Tyler Doggett & Daniel Stoljar (2010).

If we accept that there is a genuine sense of “imagination”- “modal imagination”- in which one can “imagine an arrangement of objects and properties” in a scenario, in addition to perceptual imagination and sensory or first-person-subjective imagination- imagining *what S is like* from the perspective of a subject in scenario S- it is irrelevant whether our ordinary methods of imagination will produce a positive conception of a zombie.

Marcus (2004) argues that, when we take ourselves to conceive of zombies, we are not actually forming a positive conception of a non-phenomenally-conscious creature physically identical to ourselves- rather, we are positively (perceptually) imagining something physically identical to a human being and then *refraining from imagining* (its) conscious experience. Since “refraining from imagining” does not suffice for forming a positive conception of a scenario, and the only way to conceive of an absence of subjective experience is to refrain from imagining subjective experience, we cannot positively conceive of zombies. I believe Marcus correctly diagnoses the way in which most people imagine zombies. And I agree that we cannot positively conceive of an absence of subjective experience, but only because to do so- as Marcus himself says- we would need to “sympathetically imagine” the zombie’s experience (to use Nagel’s term), to “put ourselves in a state that resembles it mentally”. And this is impossible, since there is no state that resembles the zombie’s subjective state mentally. There is no subjective state for it to mentally resemble! As Marcus says,

The problem for this thought-experiment is not that there is something imaginable, only we can’t quite conjure it up. The problem is that there is nothing to be imagined. To ‘imagine’ creatures that are objectively identical to us with all subjectivity removed is neither an act of third-person imagining, nor an act of first-person imagining. No, to ‘imagine’ a zombie is not really to imagine at all. (2004: 483)

Unfortunately, neither “sympathetic imagination” nor “first-person imagination” are “modal imagination”, and though Marcus claims that, “it is crucial for the argument in favor of the possibility of zombies that the conceivability of zombies be in part a matter of first-person imagining” (ibid: 483), this is not so. It is crucial to “modally imagining” a scenario that we can imagine certain properties (in this case, mental properties) themselves having certain properties: to modally imagine a zombie-world we must imagine that the properties of consciousness (or the properties that constitute conscious experience) are (or could be) such that creatures can share all our physical properties but lack our conscious experience. And conceiving of mental properties as bearing- or not bearing- certain relations to other properties (e.g., physical properties) does not require any first-person imagining but (if the conceivability is a priori) mere armchair conceptual analysis.

Marcus’s interpretation of Chalmers’s notion of “modal imagining” is particularly enlightening, as it highlights the notion’s deceptive simplicity.

according to Chalmers, positive imagining need not be perceptual. I can imagine what is ‘beyond the scale of perception: for example, molecules of H₂O or Germany winning the Second World War’ [Ibid., 151]. In such cases, there are no mental images, instead ‘we have an intuition of (or as of) a *world* in which S, or at least of (or as of) a situation in which S, where a situation is (roughly) a configuration of objects and properties within a world’ [Ibid., 151]. In the case of *modal* imagining, then, the mediating object is an *intuition* that represents the possibility in a way analogous to that of the image in the case of perceptual imagining. (2004, 479)

The footnote following Marcus’s quote continues, “I don’t see exactly how an intuition could be an object analogous to an image; but I won’t digress by disputing that here” (ibid: 479). In failing to see how an intuition is analogous to an image, Marcus is on to something: images and intuitions may both be a sort of mental representation, but beyond this there is no *illuminating* analogy between the two. The terms “intuition”, “image”, and

“imagination” all serve to make the extraordinary task of “modally imagining” a scenario- attempting to precisify the content of certain concepts (such as *consciousness*, *conscious experience*, *physicality*, *etc.*) and judge whether certain properties could bear certain relations to other properties- look like the commonplace act of imagining a character in a scene. To “positively conceive of a zombie” is a much more demanding chore than it first appears.

David Chalmers considers consciousness to be “at the same time the most familiar *and the most mysterious*” phenomenon in our lives (1996: 3). I think something similar should be said about *the zombie* in philosophy of mind. It is one of the most familiar characters in philosophers’ fiction and at the same time one of the most mysterious. Wrapping one’s head around the near-mindless mind of a zombie, attempting to construct an account of these fictional characters in explicit detail, is a surprisingly difficult task. And all the more so given the zombie’s apparent simplicity. Imagining a philosophical zombie seems to be a two-step process: first, imagine an ordinary person; second, remove conscious experience. There it is, a philosophical zombie. Of course, neither step is actually simple, and the “one property” that stands between ourselves and zombiehood- conscious experience- could hardly be more complex.

I do not fully trust that I can properly “modally imagine” a scenario that verifies the zombie hypothesis, that I can properly conceive of the properties of conscious experience being related to physical and functional properties in such a way that they could “come apart” as they supposedly do in a zombie-mind. But, in principle, whether or not S is *ideally positively conceivable* does not depend on our fallible powers of imagination. An ordinary epistemic agent’s conceivability judgments are “tied to [that] subject’s contingent cognitive limitations” (2010: 143), but genuine conceivability is not dictated by our limited cognitive

capacities. “Ideal positive conceivability” is the gold standard for conceivability- the kind of conceivability intended in the CP principle, and “ideal conceivability abstracts away from those limitations” (ibid: 143). In “modally imagining” a scenario, we are attempting to determine whether *an ideal conceiver* would judge that certain properties could bear certain relations to other properties.⁵⁰ Chalmers writes,

...if we are looking for a notion of conceivability such that conceivability tracks possibility perfectly, we must focus on ideal conceivability. In this sense conceivability is not a merely psychological notion; it is a *rational* notion, in much the same way that a priority and rational entailment are rational notions. If there is to be a plausible epistemic/modal bridge, it will be a bridge between the rational and modal domains. (2002: 160 emphasis original)

To make a legitimate inference from conceivability to metaphysical possibility, *conceivability* must be a rational notion and a property a scenario itself- it cannot be *subjective*, dependent on my (or our) fallible powers of “modal imagination”. Whether a scenario is ideally positively conceivable depends only on it being *possible in principle* to coherently modally imagine a scenario that verifies the hypothesis, *possible* to fill in arbitrary details about that scenario, etc., whether or not we can achieve that feat in practice.

The “common” sense among philosophers is that the zombie-scenario (thus the zombie-mind) is conceptually coherent. Property dualists and most physicalists alike believe they can “coherently modally imagine” zombies and zombie-worlds without difficulty, that they can “fill in arbitrary details in the imagined [zombie] situation such that no contradiction reveals itself” (Chalmers, 2010: 145). Most would agree that their positive conception of the

⁵⁰ Below we will look at the possibility of explaining the notion of ideal positive conceivability in terms of undefeatable justification rather than an ideal reasoner (I will argue that any explanation of conceivability- ideal or otherwise- will involve an epistemic agent).

zombie-scenario withstands the test for *ideal conceivability*,⁵¹ a point crucial for accepting premise 1 of the conceivability argument, since no “variety” of conceivability short of *ideal conceivability* qualifies as *genuine* conceivability as the property dualist intends it to be understood in the conceivability argument and interpreted in the CP principle.

Very few philosophers openly argue that this kind of mind is not ideally conceivable,⁵² and I find myself in the small minority of those who have considered this question and come to the conclusion that this kind of mind- the “zombie-mind”- fails the test for conceivability. Though at first glance we seem to have no problem conceiving of these creatures- telling stories about zombie-life, wondering how you could know whether I am a zombie, wondering if a zombie could know that he was a zombie, etc., I believe it is the very *ease* with which we can describe a zombie-world (they are so like our own!) that obscures the scenario’s inner incoherence. And regardless of whether his argument addresses the conceivability argument as Chalmers poses it, Marcus’s main point is significant: it is at least *harder than it seems* to positively conceive of the zombie-scenario.⁵³

To put the point in terms of “modal imagination”: there is no conceptually coherent way in which the relevant “objects and properties” could be arranged in order to create a scenario that verifies the conceivability of zombies. When you believe you are imagining a zombie-world, what you *think* your imagination represents cannot be what you have *actually*

⁵¹ Some physicalists might hesitate to grant such a strong form of conceivability.

⁵² Robert Kirk (1974, 2008), Daniel Dennett (1995, 1999, 2005, 2013), Sydney Shoemaker (1975), Nigel Thomas (1998), David Braddon-Mitchell (2003), Allin Botterell (2001), Keith Frankish (2007) Richard Brown (2010), as well as Marcus (2004) are among a stalwart few who have argued against the conceivability of such a mind.

⁵³ To that end I agree with Marcus, and I believe that a closer look at the “philosophical zombie” will reveal that “it” is not “there” at all.

conceived; at the end of the day, zombies, zombie-worlds, zombie minds, are simply *inconceivable*.

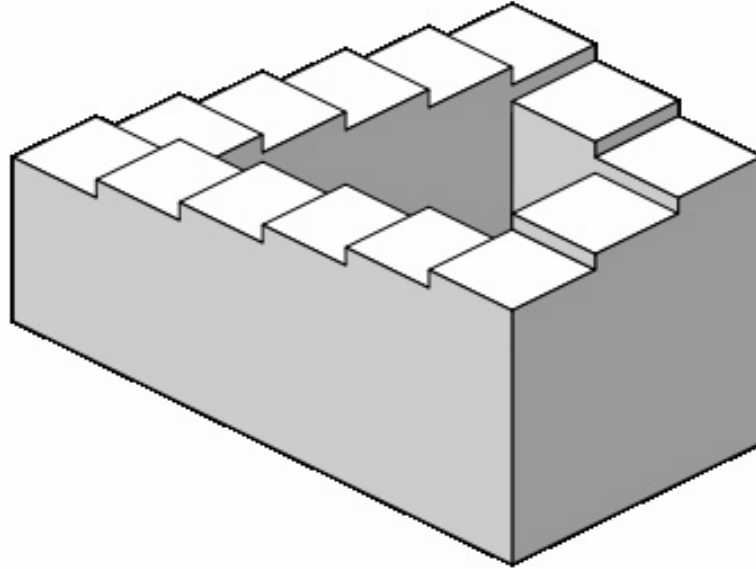
(N.B. if you are not convinced that zombies are *ideally* conceivable, or you doubt whether it is possible to *ideally positively* conceive of the zombie-scenario (rather than *prima facie* positively conceive of zombies), then- officially- so far as the conceivability argument is concerned- *you simply have doubts about the conceivability of zombies.*)

2.2 Disentangling the Conceivable from the Possible from the Actual but Incredible

2.2a Inconceivable Objects

2.2a1 Escher's Stairs

At first glance there seems to be nothing “wrong” with the picture below- Roger Penrose’s rendering of an image he conjured while attending a lecture by M. C. Escher. The second glance exposes the intentional “error”: his simple drawing depicts an impossible object- a finite ring of endlessly ascending stairs.



Escher's stairs are impossible to build- no such object can exist in ordinary three-dimensional space.⁵⁴ But is the object *conceivable*? It seems obvious that it is- at least *prima facie*. With this *image* on the page it seems you can see the ring of stairs. And you could give an articulate description of the staircase as well, something like, "It is a finite set of stairs in a closed loop; every stair connects to the one before it and behind it and each step is a step up."

To positively conceive of Escher's stairs, it is not sufficient to recall the *image* on the page or mentally re-draw Escher's picture; positively conceiving of the stairs would require mentally- or conceptually- assembling a closed loop of finite and ever-ascending stairs from the rules of geometry and the relevant concepts: *staircase*, *ascension*, *infinity*, etc. If you try to mentally create Escher's stairs in detail, to "modally imagine" Escher's stairs, your confidence in their *conceivability* may begin to wane. The more concerted your effort to complete the staircase in detail- the more detail you add to this imaginary object- the more

⁵⁴ You can construct models that *seem* to be such stairs, but only if seen from the proper angle, like the sculpture below.

the supposed object of your imagination collapses in conceptual contradictions. You cannot “fill in” arbitrary details of the scenario, of the object, without encountering a conceptual contradiction. Failing to *positively conceive* of such a staircase indicates its conceptual incoherence.

Once you spot the conceptual contradictions- recognize the incoherence of an impossible object- the identity of *what* it was that you were conceiving when, at first pass, you believed “it” was conceivable becomes philosophically puzzling. Looking back at the drawing again, the conceptually crumbled object reappears; there “it” is, and it seems so easily, so *obviously*, conceivable.

2.2a2 Penrose’s Triangle



Impossible Triangle sculpture, assembled by Brian MacKay & Ahmad Abas, located in Claisebrook Square, East Perth, Perth, Western Australia. Photo by Bjørn Christian Tørrissen.

Penrose’s Triangle has been described as "impossibility in its purest form"; though a visual (and conceptual) illusion, viewing the sculptures from the proper position brings the

Impossible Triangle to life- it seems *real* (and in that case certainly conceivable). One sits in the middle of an Australian roundabout; a shiny aluminum *impossibility*, thirteen and one half meters tall.

The Claisebrook sculpture is brilliantly composed. The size of the sculpture makes it nearly impossible to inspect all three beams simultaneously from one point of view, making the illusion all the most robust. Under the right conditions- if the object were new to you, and you viewed it from just the right angle- seeing this sculpture could challenge your concept of *triangularity*, perhaps prompt you to momentarily reevaluate your grasp of three-dimensional Euclidian geometry. But as you examine the aluminum beams composing the piece more closely, notice that the bars simultaneously move toward you and recede, you see that the corners of the “triangle” connect in ways impossible for a three-dimensional object. Take a drive around the “triangle”, view it from different angles, examine it from different perspectives, and you discover “it” wasn’t there after all.



Jennifer Matey (2012) gives an illuminating description of our perception of impossible objects- a definition that captures both the *reality* of our perceptual experience and the *impossibility* of seeing what we think we see. Impossible objects, she writes, are “experienced in perceptual character as having geometrical properties that no physically real object can have” (1). The apparent conceivability of inconceivable objects might be

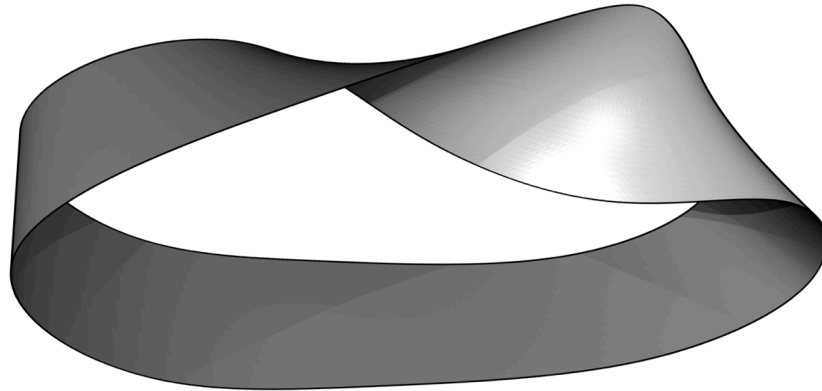
described in a similar manner. “Inconceivable objects”, or stories, or scenarios, are those that are experienced in *imagination* as having properties that no *conceptually coherent* object (or scenario) can have. Just as there can be sensory hallucinations, well-crafted arguments- or stories- or thought-experiments- can induce *conceptual hallucinations*. If “seeing” the Claisebrook sculpture led you to believe that some object could have geometric properties *like that-* that *triangularity* allowed for an object with such properties- the optical illusion induced a very real conceptual hallucination. In visual hallucinations we “see” something we really didn’t see.⁵⁵ In conceptual hallucination we imagine something we really can’t conceive. Something- some theory, argument, or thought-experiment- has made us believe we were conceiving what is really inconceivable.

Just as an M.C. Escher drawing is not invisible, philosophical zombies are not indescribable. But what Escher leads you to believe you’re seeing *isn’t actually there on the page*; in the same manner, I claim, philosophical zombies *aren’t actually “there” in the story*. Creatures with “minds” such as these- all cognition and no consciousness- are conceptually incoherent. Their description may be intelligible though the zombie-scenario, the zombie-mind itself, is inconceivable.⁵⁶

⁵⁵ How you describe your visual experience in this example- whether you visually experienced *an impossible triangle* or something else- will depend on your theory of perception. But if “see” is a success term then you certainly didn’t see an impossible object.

⁵⁶ Below I will introduce a stipulated a usage of “intelligible,” one that divorces it from the closely associated term “conceivable”, a distinction that will significantly improve the state of the debate over conceivability arguments and modal judgments.

2.2a3 An Incredible Actual Object- The Mobius Strip



Escher's Staircase and Penrose's Triangle are paradigm impossible objects, and the status of the zombie is open to debate. The Mobius Strip, by contrast, is a perfectly real but incredible object. That a three-dimensional object could have a surface with only one side (and one boundary point) is incredible, nearly unbelievable. Many of us were introduced to this amazing figure as school children, constructing one from strips of paper and tape. The Mobius Strip can create the perfect teaching moment- students who might have found mathematics and basic geometry a bit dull are confronted with a strange loop- one they've created themselves- lying right in their hands, and challenging what they thought they knew about the limits of geometric shapes. It shows them just how much they have left to learn about geometry, where our "folk-theory" of geometry runs out and matters become more complicated.

The Mobius Strip may have seemed conceptually incoherent if you had merely compared its definition with what you thought you knew about three-dimensional space. Holding and examining the *prima facie* inconceivable object in your own hands proves that

the shape is possible and the object quite real. The Mobius strip is incredible but real, talking dogs are nomologically impossible but conceivable, the existence of water without H₂O is conceivable (in a sense) but metaphysically impossible, and Escher's stairs and Penrose's Triangle are both impossible and inconceivable. How do we determine the zombies' place on this list? Do they, like the Mobius Strip, seem fantastical but in fact show us how much we have to learn about consciousness and the mind? Or are zombie-stories more like the Claisebrook Triangle Sculpture? When we listen to zombie stories, or entertain the zombie hypothesis, and we find zombies so easily intelligible that we believe they are conceivable, are we falling into the grip of a conceptual hallucination?

2.3 Possible, Conceivable, or *Merely Intelligible*?

2.3a “Intelligibility”: a Neglected Piece of the Modal Puzzle

When I claim that zombies are inconceivable I am making a sharp distinction between the *inconceivable* and the *unintelligible*; this is a stipulated use of “intelligible”, but I think the discussion of zombies and conceivability arguments would be well served by introducing a distinct notion of “intelligibility” in the vocabulary of modality, alongside “conceivability” and “possibility”, which applies expressly to the *subjective* aspect of modal judgments. *Possibility* and *impossibility* are features of worlds; *conceivability* and *inconceivability* are features of scenarios; *intelligibility* and *unintelligibility*, as I intend the terms to be used, are relations between subjects and “expressions”, where expressions are the particular linguistic terms with which a scenario is articulated.

I propose we treat “intelligibility” as a relation that holds between a subject and a particular linguistic expression used to *articulate* a scenario. There is no question whether

different linguistic expressions can be used to articulate the same scenario. It is happening all over the world as you read this sentence. A German-speaker is pondering a scenario in which *der schnee ist weiss* while an English-speaker asks herself whether snow is white. The English and German-speakers are evaluating the same scenario under two distinct expressions; though 'Der schnee ist weiss' may be *unintelligible* to the monolingual English-speaker, that is, these phonemes may be gibberish *for her*, the scenario it articulates is conceptually coherent.

To see the need for an explicitly subjective notion of *intelligibility*, distinct from “conceivability” as Chalmers uses the term, consider Chalmers’s very brief account of the intuitive evidence for the conceivability of zombies:

the question is whether the notion of a zombie is conceptually coherent. The mere intelligibility of the notion is enough to establish the conclusion.
(1996: 96)

Taken in or out of context, this claim is simply false. Conceptual coherence is always to be understood as a rational notion, but “intelligibility” here must be a subjective notion (similar to my suggested usage of “intelligibility”), and it refers to our subjective psychological relation to the zombie-scenarios described in zombie-literature. If conceptual coherence and intelligibility were both rational notions, there would be no sense in arguing for the conceivability of zombies by pointing to the intelligibility of zombie-scenarios. The intelligibility of those scenarios would be challenged on the same grounds as their conceivability, and neither would support the other.

When the subjectivity of “intelligibility” is made transparent, it is obvious that merely finding a particular description of a scenario intelligible does not establish that the scenario is (objectively) conceptually coherent. The intelligibility of the description is simply *what*

allows us to evaluate the scenario for conceptual coherence. If we found the description of a scenario *unintelligible* we would not be in a position to determine whether the scenario itself was conceivable or inconceivable. The intelligibility of the zombie-scenario does not establish its conceivability; it merely invites the *zombie* concept in for further scrutiny. Unless we divorce these two notions, any scenario we examine for conceptual coherence will automatically qualify as coherent simply in virtue of the fact that we understand the concepts involved (and their articulation) well enough to engage with the description.

Failing to note the crucial distinction between the subjective intelligibility of a scenario and its objective conceivability eliminates any hope for a fruitful discussion of the conceivability argument. It is just this sort of failure that has driven the debate over “philosophical zombies” into argumentative chaos and purported stalemates. We might still allow that the “ideal conceivability” of a scenario has implications for the possibility of the world it represents,⁵⁷ but we should deny that the character of the relation between a subject and a particular expression or description of a scenario has direct implications for the conceivability of that scenario itself. Instances of intelligibility without conceptual coherence are legion.⁵⁸ We discussed two examples above: Escher’s stairs and Penrose’s Triangle, the “inconceivable objects”. The squared circle is a third. The intelligibility of the concept “being a squared circle” is as clear as its conceptual incoherence. We can give a well-formed definition of the (contradictory) properties of a squared circle- thus the expression is intelligible- and we can use that definition to show that the notion of a squared circle is

⁵⁷ “Type B” physicalists would take issue here, but we will address this later.

⁵⁸ If intelligibility is a subjective notion, there will be conceptually coherent but unintelligible scenarios as well. We will discuss possible cases below.

conceptually incoherent- thus inconceivable. It is the very intelligibility of the notion “squared circle” that allows us to prove its conceptual incoherence.

Interpreted in this way, zombie stories needn’t be *unintelligible* for the notion of philosophical zombies to be *inconceivable*. When I claim that Escher's Stairs, Penrose’s Triangle, and philosophical zombies are inconceivable I am *not* at the same time saying that these objects or scenarios are unintelligible. Though these two types of failure often travel together, I will use them as distinct notions with independent criteria for satisfaction. The conceivability of a scenario and the intelligibility of its expression are decided by different criteria and will diverge depending on the particular expression of that scenario and the subject who purportedly understands- or fails to understand- the expression.⁵⁹

2.3a1 Intelligibility, Conceivability, and Conceptual Disagreements

A conceivability argument must start with some fairly specific, *prima facie* plausible, account of the content of a concept; the argument itself is intended to lend support to the initial (or proposed) account of this content (or provide reasons for rejecting it). Even though the *referent* of a concept may ineffable (such as the referent of *phenomenal experience* or *consciousness*), the *content* of the concept must be clearly established or there is no foundation on which to create, or evaluate, a conceivability argument.

⁵⁹ In drawing a distinction between “intelligibility” and “conceivability” here, I am interpreting the notion of conceivability as Chalmers does, in the way it purportedly operates in the conceivability argument (as a rational notion, on a par with a priority and entailment). Later I will argue that no notion of conceivability can abstract away from the conceiver altogether, and that all conceivers have limited epistemic access to the content of their concepts, at least potentially, in a way that they could not detect. If this is correct, one might argue that- at the end of the day- conceivability and intelligibility, when properly construed, will be decided on the same criteria. Therefore, when conceivability is *properly* construed, I have no objection to marrying the two notions which I am here arguing to divorce.

Latent disagreements over the content of the concept *philosophical zombie* have spawned vast and only mildly productive literature devoted to zombiehood. Some of the most vociferous debate over the nature of consciousness stems from whether the *actual* “reference-fixer” for the concept *consciousness* is something like “having a phenomenal feel”, or whether the primary intension of consciousness is better represented as “there being something it *seems* like to a subject”, with the emphasis on ‘seeming’ rather than ‘phenomenal feeling’.

For a fanciful example of conceptual disagreement, consider the prospect of “tele-transportation” and the concept of *personal identity*. A ‘Map-and-Send Machine’, the “transporter”, maps your molecules as it dissolves your body (possible in principle, we can assume), and sends that information- your molecular mapping- to a “receiver” which builds you anew out of fresh particles. Is tele-*transportation* conceivable? The details of the scenario are not under dispute. In tele-transportation, all that “travels” between transporter and receiver is information- your molecular map. If you hold that personal identity is a matter of *form*, and that the particular material out of which you are composed is irrelevant to what makes you *you*, tele-transportation will be conceptually coherent, and you can tele-travel with impunity.

Alternatively, you might believe that personal identity essentially involves some kind of continuity of physical matter over time- one that allows for certain physical changes: cellular death and birth, the loss of a limb, etc. Since the “you” that emerges from the “receiver” has no particle in common with the “you” who was mapped, the ‘Map-and-Send Machine’ will not have *transported* you but merely dissolved you to death. On the second

view, “tele-transportation” would *not be transportation*, thus tele-transportation scenarios would be conceptually incoherent.

Whether a tele-transportation scenario is conceivable or inconceivable, conceptually coherent or incoherent, is a matter of the content one ascribes to the concept of personal identity. Disagreement over this content will lead to conflicting conceivability-judgments regarding the tele-transportation scenario. Those for whom personal identity is a matter of *matter* will conclude that tele-transportation is conceptually incoherent. Alternatively, those judging the conceivability of tele-transportation from within a conceptual framework on which matter is immaterial to personal identity might find this form of travel perfectly conceivable.

Substantive conceptual disagreements can be found between deeply incommensurate conceptual frameworks. Take Causal Structuralism, (something akin to) Russelian Monism, and the identity conditions of fundamental properties across possible worlds.⁶⁰ According to (Hawthorne’s (2001) construal of) Causal Structuralism, “there is, for each fundamental property, a causal profile that constitutes the individual essence of a property” (2001: 362). According to Russelian Monism, fundamental properties play their causal/functional roles only contingently, such that the following scenario S_m , is conceivable.

S_m : Fundamental property p plays the mass-role in the actual world while, in a different possible world, fundamental property q plays the mass-role. Property p may exist in the counterfactual “ q -mass-role” world but have a different causal/functional role, or p might not exist at all in the “ q -mass-role” world.

⁶⁰ I was pleased to find a similar example involving Russelian Monism in Chalmers (2010). The term “sch-mass”- and my general description of the position- are lifted directly from his account.

According to the Russelian Monist, only property p would be “mass”; fundamental property q would not be “mass” but “sch–mass”. The fundamental properties p and q would be distinct properties sharing their causal/functional role. According to the Causal Structuralist (as I construe her), scenario S_m is internally inconsistent: that there could be two *distinct* fundamental properties playing *identical causal/functional roles* in their respective worlds is conceptually incoherent. In the “ideal conceivability” sense of conceivable, S_m is inconceivable.

Whether S_m is intelligible is an entirely different question. Assume, as I suggest, that intelligibility is not a property of a scenario itself but rather a property of the *relation* between a subject and the expression with which the scenario is articulated. If intelligibility is a *subjective* notion, no expression of a scenario will be intelligible or unintelligible *simpliciter*. Recall the earlier example in which the expression, 'der schnee ist weiss,' is both intelligible and unintelligible; that is, the expression is intelligible to a German-speaker but unintelligible to a monolingual English-speaker. And a subject may find a scenario intelligible when they believe it to be *conceptually incoherent*. The Monist’s expression of scenario S_m above is doubtless *intelligible* to the Causal Structuralist, and the Structuralist’s reaction to S_m - her denial of its conceptual coherence- is most likely intelligible to the Russelian Monist.

To say that you or I find an expression of a scenario *intelligible* does not imply that we understand *how* the concepts that compose the scenario could combine in such a way as to be conceptually coherent. To say that the expression “x is a square-circle” is intelligible does not imply that we understand how square and circular could combine in such a way that a plane figure could possess these two properties simultaneously. Neither does it imply that

“x is a square circle” is *prima facie* conceivable- anyone with a passing familiarity with squareness and circularity will immediately recognize the inconceivability of a square circle.

For a dyed-in-the-wool Causal Structuralist, S_m may not be *prima facie* conceivable, either. She may say that she finds S_m *intelligible*, but this may only mean that she sees nothing defective in the grammar or individual use of terms in S_m , and- through some analogy- she recognizes how *some scenario similar to S_m* could be true *if* that scenario involved something other than fundamental properties (cats, perhaps). But it does not imply that she understands *how* fundamental properties q and p could have *identical causal/functional profiles* and yet be *distinct fundamental properties*. Though she may find S_m intelligible, its intelligibility may not even give the Structuralist conceptual pause, and does not imply that she finds it *prima facie* conceivable.

It is possible that the Structuralist is so deeply dyed that she actually finds S_m *unintelligible*. Anti-materialists (and certain physicalists, most notably Joseph Levine) claim that any explanation of conscious experience in terms of physical processes would be unintelligible. Many “a priori” physicalists claim that the idea of creatures who are molecule-for-molecule duplicates of human beings but lacks human *consciousness* is (at least inconceivable and often) unintelligible as well.⁶¹

⁶¹ In other work (in progress) I consider the possibility that zombie-talk, though entirely systematic- both from an objective point of view and from the zombies’ point of view- is nevertheless meaningless. Some claim that the idea that zombie-talk is meaningless is absolutely *unintelligible*, that *it makes no sense* to say that a creatures’ “talk” is unfailingly “apparently” systematic and yet meaningless. I agree that this interpretation of zombie talk is by and large *unintelligible*, that it makes no sense to us psychologically and is all but impossible to wrap one’s head around. I for one cannot wrap my head around it- whenever I imagine such a scenario I can’t help but *push* meaning into the words. But I consider that the “meaningless” interpretation may be conceptually coherent nonetheless (though I argue to reject that interpretation).

In my own case, I know that I cannot wrap my head around a description of sets of physical dimensions that are themselves wound around each other as in a “string”- I understand that my faculties of imagination are limited and work within three, perhaps four, dimensions. If visualizing beyond four dimensions is *beyond* my current cognitive abilities and entertaining the idea of *dimensions* as *shaped* would stretch my cognitive capacities beyond their limits, string theory is *unintelligible* to me, though it may well be conceptually coherent (as I assume it is).

But just as intelligibility is a subjective notion, it can also be a matter of degree: a particular expression of a scenario may become more (or less) intelligible to a particular subject for different reasons. The monolingual English-speaker may become a German student, for example, and 'der schnee ist weiss' may become more intelligible as she becomes more adept at using the German language. I may become more enlightened in the ways of theoretical physics, and string theory may become more intelligible to me than it was before. And the Structuralist- or anti-materialist, or physicalist- may begin to question parts of her conceptual framework, revising her understanding of certain concepts; she may find formerly unintelligible descriptions of scenarios increasingly intelligible, opening conceptual paths through which she may come to judge the scenarios themselves to be at least *prima facie*, and perhaps ideally, conceivable.

Despite Chalmers's claim to the contrary, the mere intelligibility of zombie-stories does not guarantee their conceptual coherence. Intelligibility is subjective and fluid, while conceptual coherence is objective and fixed. Conceivability judgments reflect only the content of concepts and their relation to each other. Intelligibility may be a guide to conceivability, but as the former is subjective and a matter of degree while the latter is not,

the former cannot straightforwardly imply the latter. How ought we distinguish between the incredible but real, the impossible but conceivable, and the genuinely inconceivable, if any one of these scenarios may conflict with the others and all (or none) may be intelligible to us? Even *expert* concept users can “believe the inconceivable” when they are confronted with a scenario that is both *suitably complex* and *highly intelligible*. A compelling philosophical thought-experiment has both these qualities. The intelligibility of its description obscures the scenario’s complexity, and its complexity in turn hides its conceptual incoherence. There is nothing devious about such a thought-experiment or conceivability argument; the authors and proponents of such arguments are just as much “taken in” by the scenario’s intelligibility as the dissenters.

We need to know something about stairs and triangles and three-dimensional Euclidean space before we can get a handle on Penrose’s and Escher’s perplexing “objects”. In the geometrical cases we have something to keep us grounded: images to scrutinize, objects to pick up and run our fingers around or view from different angles, and rules of geometry, etc., to which we can defer. In the case of string theory we have mathematics and theoretical physics.

In the case of the zombie we have no such luxuries. No authoritative laws to turn to, no axioms or established “truths” of consciousness of the kind we find in mathematics. And no amount of poking at a mass of gray matter will tell us whether there is a “mind inside” or what its experiences are *like*. We have only concepts to consult- *consciousness, mind, experience...* and the meaning of these concepts themselves is widely disputed. So how do we determine the zombie’s place in the logical space of actual/possible/inconceivable objects? We find ourselves with the task of determining what kinds of minds are conceivable

and inconceivable, possible or impossible, without a consensus on the concept of *conscious experience* itself. “Having a conscious experience” may mean “having a phenomenal feel,” or it may mean “*seeming* to a subject to have a “phenomenal” feel;” two radically different accounts of the content of the concept of *consciousness* resulting in two mutually exclusive views on the conceivability of zombie-minds.

It is unlikely that such a hotly debated concept as *consciousness* will be “well behaved” enough to allow us to pin down its content entirely- any account of the content of *consciousness* will be controversial.⁶² But since the conceivability of zombies is determined solely on grounds of the content of the concepts out of which the zombie-scenario is constructed, we need an explicit account of the content of the concepts constructing the scenario- an account more precise than “the common sense notion of S” or “our pretheoretical understanding of S”. If parties are coming to the debate with implicit disagreement over the meaning of concepts crucial to the argument, their verdict on whether S is conceivable will fail to show anything about the coherence or incoherence of the scenario itself. There will not be one scenario under discussion- there will be as many scenarios as there are interpretations of the content of the concepts use to describe it.⁶³

The first step, then, in evaluating a conceivability argument is *determining the content of the concepts as the author intends*. Certain arguments might succeed only if one accepts

⁶² It may well turn out that it is impossible to specify the (metaphorically) “microphysical” details of the zombie-scenario- what we might call the “zombie-minutia”- in such a way that both dualists and physicalists could still agree zombies are genuinely conceivable.

⁶³ If one of the “virtues” of the conceivability argument is that it verges on the edge of being a deductive argument, one of the “vices” of conceivability arguments in general is that anyone who objects to the proposed content of the relevant concepts is unlikely to be convinced of the argument’s conclusion. But since this is a “shortcoming” of all nearly all philosophical argument, it is not a particular failing of the conceivability argument but simply more pronounced.

the particular meaning or content the author assigns to the essential concepts (as tele-*transportation* is conceptually incoherent if personal identity is a matter of material continuity, but perfectly conceivable if a *soul* or *form* is the essential component of personhood). Without an understanding of the content of these concepts *as the author intends*, you can neither evaluate the argument for consistency nor argue that the assigned content of the concepts does not accurately represent their meaning (or that the author has skewed the concepts order to arrive at a certain conclusion, etc.).

The second step is to determine *the content to which the author is actually committed*. The key to a conceivability argument's undoing is a discrepancy between the *intended* and the *actual* content of the concepts employed in the argument- not "actual content" as in what the concept *really* means in the *real* world (it's *real* "reference-fixers") but the content of the concepts as determined by their use in the argument under consideration. Depending on the discrepancy, this might either lead us to question the plausibility of the premises or it could expose a conceptual inconsistency, undermining the validity of the conceivability argument altogether.

3. The Dualist's Essential Task: Defending our Epistemic Situation from the Zombie-Occupation

3.1 Zombies, the Knowledge Argument, and the "Phenomenal Concept Strategy"

The knowledge argument is often thought of as the epistemic arm of the conceivability argument, the two arguments purportedly two sides of the same anti-materialist coin. While the conceivability argument argues for the existence of non-physical properties from the conceivability of zombies, the knowledge argument argues for the existence of non-physical facts from the apparent inexplicability of facts about what it's like to have a conscious experience from complete knowledge of the physical facts alone. Mary the Color Scientist is the protagonist of its best known thought-experiment, and her creator, Frank Jackson, describes her situation as follows:

Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. (1986: 291)

Mary's scientific color-knowledge is exhaustive. She knows all the scientific facts about color and human color experience that there are to know; that is, all the color-facts one

can learn without actually having color experience. In the thirty years following her inception, Mary's story has been extensively revised. Subsequent versions have expanded the ways in which she is prohibited from having color experience; her skin was once painted shades of grey, she was forbidden from rubbing her eyes so as not to produce colorful pressure phosphine, she had been prohibited from dreaming in color, etc. Eventually she was simply re-created as colorblind from birth. All these roads lead to the same Super Scientist: Mary is supposedly functionally identical to a normal human being except that she has *not yet* experienced seeing color. And she knows all the scientific facts about color and human color experience- perhaps all the scientific facts, full stop. On some accounts, Mary is simply *scientifically omniscient*. As Jackson continues,

If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies...

It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning-she will not say, "ho, hum". Hence physicalism is false. (ibid: 291)

The day that Mary leaves her room- or her color-blindness is cured- and steps out into the world of color, upon seeing her first red object she supposedly has a thought, something like "oh, so *that's* what it's like to see red!" As Jackson says, she will not say "ho, hum". Her reaction to seeing red is evidence that she learns something upon her first red-experience- she learns *what it's like* to have an experience of red.

Does this imply the falsity of physicalism?

The knowledge argument is taken by anti-physicalists to show that there are more facts in the world than physical facts; there are *phenomenal facts*, facts about the qualitative

aspect of experience, about *what it's like* to see red, to hear a symphony, to ride a motorcycle in the rain, that cannot be captured by stating all the physical facts about the world, the facts that Mary knew prior to her release. And these non-physical facts supposedly have ontological implications: the non-physical *facts* that Mary learns are facts about the non-physical *properties* that exist in our world. Thus Mary's new knowledge implies that materialism is false of our world.

Physicalists deny that Mary learns a new non-physical *fact* when she leaves her black and white room, but agree that something cognitively significant happens to Mary upon her release. According to a posteriori physicalists, Mary acquires a new *phenomenal concept*, the phenomenal concept "red-experience" (the concept which she expresses in her exclamation with the term "*that*"). This new concept allows Mary to think about familiar facts- facts about color and color vision that she knew while inside her black and white room- in a new *way*. Thinking about old facts in a new way is importantly *not* a form of acquiring knowledge of *new* facts. Mary's new phenomenal concept allows her to refer to red-experience *directly*, from the subjective perspective, rather than mediated by theoretical knowledge and inference.

3.1a Using "Conceptual Isolation" to Block Entailment

For the possibility of a scenario, S, to be *entailed* by the conceivability of S, further criteria must be satisfied beyond the ideal positive primary conceivability of S. In particular, there can be *no viable alternative explanations* of S's conceivability. The difference between primary and secondary conceivability may have explained why some metaphysically impossible scenarios *seem* conceivable, but it does nothing to prevent "blockers"- *multiple*

scenarios, all ideally positively conceivable, that explain the conceivability zombies without appealing to an additional class of ontological properties, thus “blocking” the property dualist inference from conceivability to possibility.

Dualists and *a posteriori* physicalists agree that phenomenal and non-phenomenal concepts are “isolated” from one another- facts couched in phenomenal terms cannot be inferred from facts couched in entirely physical terms, and (relatedly) statements expressed in phenomenal terms cannot be translated into statements expressed in entirely physical or functional (non-phenomenal) terms without some loss of meaning. Some physicalists (applying what has come to be called “The Phenomenal Concept Strategy” (Loar, 1990)) claim that this isolation explains why Mary cannot know what it's like to see red simply in virtue of knowing all the physical facts about her world. Physical facts are facts expressed with physical concepts, and though physical concepts and phenomenal concepts may have the same *referent* they do not share their *meaning*, thus no knowledge of facts expressed with physical concepts will close the explanatory gap, or allow Mary to acquire a *phenomenal* concept. Physicalists who support the phenomenal concept strategy believe that the isolation of phenomenal concepts from physical concepts can account for the explanatory gap between physical facts and facts about conscious experience.

There is no single definitive version of the phenomenal concept strategy,⁶⁴ but all forms share two basic steps in common. The first is to claim that the psychophysical

⁶⁴ But see Stoljar (2005) for a thorough analysis of the similarity and differences between different versions of the phenomenal concept strategy. Loar (2006) was the first to call this collection of a posteriori physicalist responses by this name.

conditional “ $P \rightarrow Q$ ”⁶⁵ is a necessary *a posteriori* truth, and grant that it appears to be contingent.⁶⁶ The second step is to acknowledge the conceptual isolation of phenomenal concepts from physical concepts and argue that conceptual isolation accounts for both (1) the absence of any *a priori* entailment of phenomenal facts from physical facts and, relatedly, (2) the psychophysical conditional’s appearance of contingency.

3.1b The Phenomenal Concept Strategy (an Informal Formalization)

Consider the three claims below:

- a) Zombies are conceivable.
- b) Mary acquires some sort of new and nontrivial propositional knowledge upon seeing red for the first time.⁶⁷
- c) Psychophysical identity claims *appear* to be contingent, rather than necessary.

Property dualists take each of the claims (a), (b), and (c) to show that physicalism is false. An *a priori* physicalist could deny any- most likely, all- of these three claims, while his *a posteriori* physicalist counterpart will grant all three claims while maintaining that their truth does not imply the falsity of physicalism. In terms of (a), (b), and (c) above, we can roughly formalize the *a posteriori* physicalist’s appeal to phenomenal concepts as follows:

1. All phenomenal facts in world w_1 are metaphysically necessitated by the physical facts in w_1 .
2. Psychophysical identities are necessary *a posteriori* truths.

⁶⁵ P is the set of all physical facts in a world, and Q is the set of all phenomenal facts in that world.

⁶⁶ According to Papineau (2009), it simply appears *false*.

⁶⁷ There are also *a posteriori* physicalists that deny the significance of phenomenal concepts- these physicalists will reject (b). See Levin (2007: 95) for a discussion of the triviality of the knowledge that Mary gains upon seeing red for the first time.

3. Phenomenal concepts are conceptually isolated from physical concepts.
4. It follows from (2) that psychophysical conditional is true in all possible worlds $\Box(P \rightarrow P^*)$
5. (3) implies that phenomenal facts cannot be deduced *a priori* from physical facts.
6. (4) and (5) explain (or are compatible with) the truth of (a), (b) and (c) above.
7. The truth of (a), (b), and (c) is compatible with (1) the truth of physicalism.

If the phenomenal concept strategy works- if the source of the explanatory gap is simply the isolation of physical concepts from phenomenal concepts, and the existence of irreducible, non-physical, phenomenal properties is not entailed by the explanatory gap or by the conceivability of zombies simpliciter- then the nature of the relationship between our physical and functional concepts will deflate the metaphysical *significance* of the gap. Our phenomenal concepts do not *close* the gap, nor does the a posteriori physicalist intend them to. If conceptual isolation does hold between physical and phenomenal concepts we should *expect* to see an epistemic gap even though physical states are identical to phenomenal states.⁶⁸ If the phenomenal concept strategy succeeds here, the explanatory gap provides no evidence for the claim that property dualism is true of our world.

⁶⁸ Phenomenal states may also metaphysically supervene upon physical states, without remainder.

3.2 The Dualist's Argument for Epistemic Inequality: a "Dilemma" for the Physicalist

The purpose of the phenomenal concept strategy is to explain our epistemic situation regarding consciousness by appeal to the nature of our physically explicable phenomenal concepts alone. Chalmers's (2007) argument against the phenomenal concept strategy is based on the claim that one defining feature of our epistemic situation is that there are "distinctive epistemic gaps" (172) between consciousness and physical facts that exist in our world, epistemic gaps that do not exist in the zombie-world. Chalmers's criteria for "sharing an epistemic situation" are that subjects have corresponding beliefs with corresponding truth-value and corresponding epistemic status "as justified or unjustified, as cognitively significant or insignificant" (2007: 176). Despite the zombie-dualists' and zombie-a posteriori physicalists' protest to the contrary, whatever corresponding beliefs our zombie-twins have about an "epistemic gap" in their world will be either false or less justified than our own.⁶⁹ As such, zombies cannot share our epistemic situation regarding phenomenal consciousness.

The central purpose of the phenomenal concept strategy is to establish that the nature of the relationship between phenomenal concepts and non-phenomenal concepts in our cognitive economy can account for our epistemic situation. But on a physicalist account of phenomenal belief, our phenomenal beliefs and zombie "phenomenal" beliefs would be formed by the same causal or functional mechanisms, equally reliable, and as such the

⁶⁹ There is the possibility that these zombie beliefs are meaningless, though we will not entertain that option here.

justification of human and zombie phenomenal beliefs will be equally strong and their corresponding beliefs will have corresponding truth value. The fatal flaw in the phenomenal concept strategy is that any fully physical accounts of phenomenal concepts will explain the zombies' epistemic situation regarding consciousness as well as our own, making the physicalist phenomenal concept strategy insufficient for explaining why we find "distinctive" epistemic gaps in our own world, and why we can conceive of zombies, inverted-twins, and why Mary learns something when she encounters color for the first time.

Subjects may share their epistemic situation though the content of their corresponding beliefs might differ. Both physicalists and property dualists agree that, "the claim that a zombie and conscious being share their epistemic situation does not require that their beliefs have the same content... epistemic situations should be understood in topic-neutral terms," to avoid begging the question against the physicalist (2007: 177). Differences in the content of human and zombie subjects' corresponding "phenomenal" beliefs- *on its own*- is irrelevant to our comparative epistemic situations, just as the difference between the content of Oscar's and Twin Oscar's corresponding '(t)water' beliefs is irrelevant to whether they share their epistemic situation.

This is the theory behind Chalmers's much-discussed dilemma for the phenomenal concept strategy, which he described informally here, arguing that no physicalist account of phenomenal concepts could be both

...powerful enough to explain our epistemic situation with regard to consciousness, and tame enough to be explained in physical terms. That is: if the relevant features of phenomenal concepts can be explained in physical terms, the features cannot explain the explanatory gap. And if the features can explain the explanatory gap, they cannot themselves be explained in physical terms. (2007: 168)

More formally, we can present Chalmers's argument in this way: take C below to be an account of the entire set of mental features of human beings, both functional and phenomenal, focusing particularly on how we acquire and deploy phenomenal concepts.⁷⁰ Take P to be the set of all and only physical facts about a human being, or of the world as a whole, if the difference is significant. Chalmers gives the phenomenal concept strategist two options.

1. Accept that P&~C is conceivable
2. Accept that P&~C is inconceivable

In either case, the phenomenal concept strategy supposedly fails.

Consider the first horn of the dilemma. Assume P&~C is conceivable, and that physicalism is true, such that C metaphysically supervenes upon P. This supervening set, C, includes our acquisition, possession, and use of phenomenal concepts. Prima facie, if the physicalist embraces the first horn, she merely grants that the falsity of physicalism is conceivable; if zombie *creatures* are conceivable, it seems an inconsequential step to agree that one can conceive of phenomenal *concepts* having a non-physical feature as well. This modesty is common among a posteriori physicalists.

But Chalmers takes the conceivability of P&~C to imply something stronger. If P&~C is conceivable, then our phenomenal concepts, which, as concepts, are part of C, i.e., the set of all mental states are, “not tame enough to be explained in physical terms” (ibid: 168). The conceivability of P&~C supposedly implies that there is some feature or features

⁷⁰ In the context of this “dilemma”, Chalmers uses the phrase “psychological feature” to encompass all aspects of the mental, phenomenal as well as functional. This diverges from his standard use of “psychological”; ordinarily, it refers only to mental processes/states that are fully explicable in functional terms. To alleviate any confusion here, I will use the term “mental state” where he uses “psychological”, except in direct quotations.

of C, the set of all mental facts about ourselves, that P, the set of all physical facts, cannot explain. If the set of all physical facts itself contained a full explanation of the nature of our phenomenal concepts such that these phenomenal concepts gave rise to the explanatory gap, C would be a proper part of P, and P without C would be conceptually incoherent, i.e., *inconceivable*.

Per the second horn, if $P \& \sim C$ is inconceivable, there would be no conceivable difference between any aspect of our mental states and our physical duplicates' mental states, including our phenomenal concepts and phenomenal mental states. The inconceivability of $P \& \sim C$ implies that any creature who we conceive of as sharing our physical properties we necessarily conceive of as sharing our mental properties across the board; both zombies and inverters would be inconceivable, as the a priori physicalist maintains. It would be inconceivable for any phenomenal facts to differ between two worlds in which the physical facts are identical. To claim that $P \& \sim C$ is inconceivable would be to deny the epistemic gap between physical knowledge and phenomenal knowledge, which both a posteriori physicalists and dualist agree exists.

To embrace horn (2) of the dilemma would be to give up a posteriori physicalism for a position closer to a priori physicalism. As Chalmers sets up his "master argument",

...either physical duplicates that lack the key features are conceivable, or they are not. This allows us to set up a master argument against the phenomenal concepts strategy, in the form of a dilemma:

1. If $P \& \sim C$ is conceivable, then C is not physically explicable.
2. If $P \& \sim C$ is not conceivable, then C cannot explain our epistemic situation.

3. Either C is not physically explicable, or C cannot explain our epistemic situation. (2007: 174)

The “key features” here are the physicalist’s phenomenal concepts that supposedly account for the existence of the explanatory gap. There are myriad responses to this dilemma, but given the way the argument is posed, the key point of contention is the state of our epistemic situation. Chalmers argues that the phenomenal concept strategist must provide some theory of phenomenal concepts that explains why we can *conceive* of physical duplicates of ourselves who are in a different epistemic situation from our own, and that no physicalist account of phenomenal concepts has the resources to do so.

The physicalist might try to argue that since it is *merely* conceivable for our epistemic situation to differ from our zombie-twins’, given that zombies are metaphysically impossible, any epistemic inequality between ourselves and zombies would be impossible as well. Our phenomenal concepts need not explain why we can conceive of zombies in a different epistemic situation than ourselves, since no zombie would ever be in a different epistemic situation than her human twin, given that zombies would never exist.

But the property dualist has good ground for claiming that this physicalist defense of the phenomenal concept strategy would be weak. If our phenomenal concepts explain *our* epistemic situation, and it is conceivable that 1) zombies do not share our epistemic situation (that is, we can conceive of zombies as *not* sharing our epistemic situation), then it must be conceivable that 2) zombies do not share our phenomenal concepts. But so long as a concept- and our possession of that concept- is physically explicable, our zombie-twins must have that concept as well; this general rule would apply to phenomenal concepts as well as any ordinary, non-phenomenal, concepts. So our zombie-twins will possess the same concepts that- according to the phenomenal concept strategy- explain our epistemic situation;

whatever features of our phenomenal concepts that explains our epistemic situation also explains the zombie's.

His case for premise 2 of the master argument above spells this out directly:

Premise 2 says that if $P \& \sim C$ is not conceivable, then C cannot explain our epistemic situation... One can put the case [for premise 2] informally as follows:

4. If $P \& \sim C$ is not conceivable, then zombies satisfy C .
 5. Zombies do not share our epistemic situation.
 6. If zombies satisfy C but do not share our epistemic situation, then C cannot explain our epistemic situation.
-
7. If $P \& \sim C$ is not conceivable, then C cannot explain our epistemic situation.
- (2007: 176)

3.3 Dualists' and Physicalists' Differing Conceptions of Phenomenal Concepts

3.3a A Fundamental Division- Constitutional and Non-Constitutional Phenomenal Concepts

The a posteriori physicalist's phenomenal concepts have a large burden to bear. They must create an explanatory gap between physical and phenomenal facts while denying any metaphysical implications, they must account for the conceivability of zombies without allowing for their possibility, and they must explain how Mary can know all the physical and functional facts relevant to color and color experience without thereby knowing what it would be like to have a red-experience.

Most of the accounts of phenomenal concepts that are relevant to our discussion here assume some kind of representational theory of mind, by which I only mean that they take thoughts, beliefs, judgments- propositional attitudes in general- to consist of mental

representations; these mental representations refer, they sometimes have a truth value, can be justified or unjustified, and so forth. This kind of “representationalism” treats the act of thinking as the manipulation of expressions in an internal “language of thought” or *Mentalese*.⁷¹ Thoughts/beliefs are “sentences” in *Mentalese*, and concepts are the mental units out of which these “*Mentalese* sentences” are composed (roughly analogous to the relation between words and sentences in a public-language)⁷². As an overly simplistic starting-point, we may treat concepts as the constituent components of beliefs, and beliefs as “sentences” of *Mentalese*.⁷³

Within this “mental representation” framework we can sort accounts of phenomenal concepts by the general relation that phenomenal concepts bear to their referents, phenomenal properties. For any theory that assumes a language of thought, the connection between a phenomenal concept and its referent will either be some kind of causal connection, or in some way tighter-than-causal. In the second case, the referent is thought to be part of the content of the concept itself, somehow *present* in the concept, or *taken up* into it. These are “constitutional” accounts of phenomenal concepts; the referent- the phenomenal

⁷¹ Jerry Fodor (1987) arguably coined this now-common term.

⁷² The “language of thought” view of mental processing or belief is not necessarily reflect the views of the author. There are many viable accounts of belief, some of which do not endorse a “language of thought” hypothesis, others of which treat concepts as abstract entities we “grasp”, rather than mental entities we create. I mention this only to put it aside, since most of the accounts relevant to our discussion endorse a Language of Thought view of beliefs and concept possession.

⁷³ This discussion leaves aside “minimalist” or “thin” accounts of phenomenal concepts, according to which phenomenal concepts only barely qualify as concepts (if they qualify at all). On Prinz (2005), theory of phenomenal concepts, they are no more than *mental pointers*, where “mental pointers” are, “phenomenal demonstratives [that] refer to the conscious perceptual states that are made salient by mental demonstration... [they] use representations of objects in space to direct focal attention on a perceived scene. They are individuated by their causal powers” (2005: 13). Tye (2009) claims there are no phenomenal concepts in any significant sense; those that do exist carry very little, almost negligible, phenomenal information.

property- partially constitutes the concept itself. Accounts according to which phenomenal properties are not “present in” the phenomenal concepts themselves are, by contrast, “non-constitutional” accounts.

3.3b Non-Constitutional Accounts

3.3b1 Phenomenal Concepts as Pure Demonstratives

The majority of physicalist accounts of phenomenal concepts are non-constitutional (with notable exceptions that will be discussed below). Within this broad class there are, of course, finer discriminations (though no further “official” categorizations, as different theories are often placed under different headings depending on which features of the theory are being emphasized in a particular context). There is a fairly wide category of accounts that consider phenomenal concepts to be a kind of “type-demonstrative” concept, one that picks out (“that *kind* [of x]”), where x is a phenomenal property and “that *kind*” is a phenomenal concept referring to a phenomenal *type*. As with other demonstratives, type-demonstrative phenomenal concepts pick out their referent directly; they have “no reference-fixing “modes of presentation” or Kaplanian “characters” (Levin: 89). These concepts,

purport to pick out *kinds of properties* of experience from an introspective perspective... their references are determined solely by the causal and dispositional relations an individual has to her internal states that are effected by an introspective “pointing in”; that is, by the fact that she’s in causal contact with a certain property and is disposed to reidentify it on subsequent occasions. (Levin, 2007, 89)

Continuing from our previous discussion of qualia, this account of phenomenal concepts, though quite clear as an account of the concept, is neutral between the sort of “Q property” quale, qualia in the weak sense (qualia that metaphysically “supervenes upon” or are reducible to physical or functional properties) and “strong qualia” (in Block’s sense); as

she refers to ‘kinds of properties *of experiences*’, these could be higher-order properties or properties of higher-order properties. But the referent will be determined in the same way, regardless of how one defines “properties of experiences”, by the causal relations between and individual and her internal states that “point inward” at the quale.

Brian Loar’s (1990/97) “recognitional/demonstrative” theory is also representative of a type-demonstrative account.⁷⁴ On his view, phenomenal concepts are “self directed recognitional concepts” that,

have the form ‘x is one of *that* kind’; they are type-demonstratives. These demonstratives are grounded in dispositions to classify, by way of perceptual discriminations, certain objects, events, situations. (1990: §2)

As a recognitional concept, a phenomenal concept picks out its referent (e.g., the phenomenal property “what it’s like to see red”) as one of *those* properties without appealing to any theoretical or background knowledge of the property. Michael Tye’s (2003) rendering of the type-demonstrative account emphasizes its *causal* aspect.⁷⁵ As he explains, the phenomenal concept, C,

refer[s] to a phenomenal quality Q via C’s being the concept that is exercised in an introspective act of awareness by person P if, and only if, under normal conditions of introspection, C is tokened in P’s current experience because Q is tokened. (2003: 7)

3.3b2 Phenomenal Concepts as Impure Demonstratives

In contrast to the pure demonstrative model described above, there are physicalist theories of phenomenal concepts that do take phenomenal concepts to employ some variety

⁷⁴ As well as Lycan (2003), (on some readings, Block (2002), Sturgeon (2000), and, perhaps, Block (2002). Block (2007) can be read as endorsing a constitutional account.

⁷⁵ Though his revised position differs and is noted above.

of “mode of presentation” (though not a *descriptive* mode of presentation) when identifying their referent, rather than simply “pointing toward” a property in some kind of direct, inwardly-aimed, demonstrative act. Perry’s (2001) account retains to some of the spirit of pure demonstrative accounts (as Levin describes them), but on his view phenomenal concepts refer via a kind of perceptual state that “guides” the demonstrative concept to its referent. Being “guided by a perceptual state” can be seen as a mode of presentation by which a phenomenal concept picks out or identifies the phenomenal property to which it refers, though it should not be interpreted as a descriptive one. Conceptual isolation requires that phenomenal properties, qua *phenomenal*, cannot be described in non-phenomenal terms- there is no translation between the two kinds of concepts that preserves their meaning. Perry’s suggestion does not violate these constraints and is still, at core, a demonstrative account, though (per Levin’s (2007) interpretation) an embellished rather than “pure” demonstrative.⁷⁶

3.3c Constitutional accounts

Constitutional accounts of phenomenal concepts are radically different from the “demonstrative” or “recognitional” theories described above. Katlin Balog (2012) describes the constitutional account of phenomenal concepts as an account on which,

every concept token applied to current experience is constituted by a current token phenomenal experience, and—on most versions of the constitutional account—this fact is crucial in determining the reference of the concept...

⁷⁶ The difference between the two is still worthy of mention since, as Levin argues, the embellishments in Perry’s kind of demonstrative account are unnecessary and motivated by non-physicalist “intuitions that have already been explained away” (2007: 105).

Not only is it the case that a token state that realizes a token concept is also a token of the referent, but it is because the concept is so constituted that it so refers. (8)

Though “constitution” is most often taken to be a relationship that the dualist posits between phenomenal concepts and their phenomenal referents, there are widely respected physicalist constitutional theories as well.⁷⁷ Block (2007) refers to the phenomenal concepts so constituted by their phenomenal referents as “heavy duty” phenomenal concepts. On Block’s account, these phenomenal concepts are,

individuated with respect to fundamental uses that involve the actual occurrence of phenomenal properties. In these fundamental uses, an actually occurring experience is used to think about that very experience. No one could have a phenomenal concept if they could not in some way relate the concept to such fundamental uses in which the subject actually has an instance of the phenomenal quality.

...[more precisely,] an instantiation of a phenomenal property is used in the concept to pick out a phenomenal property (a type). Of course, the experience involved in the fundamental use need not be an additional experience, that is, additional to the referent. A single experience can be both the object of thought and part of the way of thinking about that object. (2007: 252-253)

There are at least two reasons one might adopt a constitutional view of phenomenal concepts. First, one might think that treating the mechanism by which a phenomenal concept connects with its referent as a causal connection would unduly separate the concept and referent into two *distinct* mental entities, creating a distance between subject and phenomenal experience that misrepresents the relation between a subject and her own conscious experience. Second, reference via the causal connections between phenomenal concept and referent is fallible in the same way that any causal form of reference is fallible. But in phenomenal reference, a subject is picking out a property that is within her own mind and of

⁷⁷ Loar (1990/1997) gestures towards a constitution view himself, and may have inspired some of the current constitution-physicalists.

which she is conscious. So even granting that subjects *might* fail in referring to their own experiences, there seems to be a significant difference between failure of reference in this case and ordinary cases of reference failure in which the referent (object or property) is not itself a part of the subject's mind (or the subject's conscious experience).

Papineau's (2002) "Quotational/Indexical" account is a paradigmatic constitutional model (though peculiar as a physicalist an account of phenomenal concepts in general). On his view, a phenomenal concept,

incorporates the things referred to, and thereby forms a compound which refers to that thing. Thus, ordinary quotation marks can be viewed as forming a frame, which, when filled by a word, yields a term for that word. Similarly, my phenomenal concepts involve a frame, which I have represented as 'the experience: ---'; and, when this frame is filled by an experience, the whole then refers to that experience. (2002: 117)

Though the language above can be read as simple metaphor (see Levin 2007), Papineau's account does treat the property itself as a part of the concept that denotes that property. This feature is retained in his revised (2007) model as well. On this later model, "phenomenal reference to an experience will *deploy an instance of that experience*, and in this sense will *use* that experience in order to *mention* it" (2007: 123, emphasis mine).⁷⁸ Even in this revised account, the role of the phenomenal property is still analogous to the function of a "quoted" word in a sentence.⁷⁹

⁷⁸ In his revised model, he omits unnecessarily problematic terms such as "filled" in "filled by an experience," as this language gave the impression that (or had the undesirable consequence that) one could not deploy the concept phenomenal red in a negative belief such as "I am not now having a phenomenal red experience". Replacing these terms with "deploy" and "use" (in addition to other clarifications) is an attempt at resolving this particular problem.

⁷⁹ As a qualification, Papineau thinks of phenomenal concepts as information-storing folders rather than words in a language of thought. Also, interestingly, his "phenomenal" concept is only incidentally phenomenal- he

3.4 The Property Dualist's Restrictions on Phenomenal Concepts

While both demonstrative and constitutional accounts of phenomenal concepts are in principle available to the physicalist, the dualist is much more constrained.

If phenomenal properties are non-physical, as the property dualist assumes, then the nature of a phenomenal concept must be such that it can denote a phenomenal (non-physical) property rather than a physical property in cases where these two types of properties may share important features in common. If property dualism is true, then our phenomenal concepts are not purely causal-demonstrative concepts. Causal-demonstrative phenomenal concepts pick out *whatever* state causes a phenomenal concept to be deployed- as such it is not equipped to differentiate between non-physical phenomenal properties and the brain states that also play a causal role in disposing a subject to deploy a particular phenomenal concept. A purely causal/demonstrative account of phenomenal concepts does not have the resources to guarantee that the phenomenal concept refers to the non-physical *phenomenal* property rather than the causally relevant brain state.

Though we will assume there is a psychophysical law connecting the brain state and non-physical phenomenal property, a concept that merely refers to the conjunction, or disjunction, of the two properties will be unacceptable as well. The dualist could not allow that the phenomenal concept be disjunctive- a concept which could only refer to the

takes it to be a “surprising implication” of his view is that there is nothing essentially *phenomenal* about phenomenal concepts (2007: 125).

conjunction, or disjunction,⁸⁰ of physical and phenomenal causes would not be a *phenomenal* concept.⁸¹ Whatever state or property is causally responsible for a subject deploying a phenomenal concept, there will be a physical property- a brain state- playing that causal role, *even if* there is a phenomenal property in addition *also* causally responsible for the concept being deployed.⁸² Since one of the disjuncts, the causally relevant brain state, is purely physical, a zombie's counterpart brain state will satisfy the reference of a disjunctive phenomenal concept just as well as a human's phenomenal state would.

There are several reasons the property dualist could not allow that a phenomenal concept refer to a conjunction of physical and phenomenal states; at the very least, it would not serve his argumentative purposes to have the most fundamental of our phenomenal concepts- the concept deployed via introspection to pick out the a phenomenal element of a subject's first personal subjective experience (the phenomenal concept from which all other phenomenal concepts descend) to refer to a conjunction of a physical and phenomenal properties. Without purely phenomenal concepts, rather than some kind of amalgamated phenomenal/physical concepts, it would be impossible to formulate an argument for property dualism. To do so, the dualist's phenomenal concept must pick out phenomenal states without at the same time picking out any non-phenomenal state, so that he can claim that the referent of *that* concept, the purely phenomenal concept, is non-physical.

⁸⁰ Hawthorn and (?) suggest that phenomenal concepts are disjunctive concepts, referring to physical properties if physicalism is true, and phenomenal properties if property dualism is true. Chalmers argues against our phenomenal concepts working in this way- for my part, I believe it is worth serious consideration.

⁸¹ Given that the physical and phenomenal causes are ontologically distinct existences (properties)

⁸² This is guaranteed by the causal closure of the physical world under the laws of physics- orthodoxy for both physicalists and dualists alike. Any dissenters are beyond the scope of our discussion here.

3.4a Pure Phenomenal Concepts and the “near Cartesian certainty” of our phenomenal beliefs

On a property dualist’s account of phenomenal concepts, not all reference can be fixed relationally or demonstratively. There must be some kind of phenomenal concept which, as Chalmers puts the point, refers to a phenomenal property, “directly, in terms of its intrinsic phenomenal nature” (2003: 225). This is the pure phenomenal concept, which “characterizes the phenomenal quality *as* the phenomenal quality that it is” (2003: 226, emphasis original), rather than as *whatever* phenomenal property stands in a particular relation to the subject or community. The pure phenomenal concept is “a substantive concept that is tied *a priori* to a specific sort of quality” (2003: 227).⁸³

The relation between pure phenomenal concepts and their referents, our “qualia”, is unique among epistemic relations—it is unmediated and stronger than an ordinary case of direct reference. In a standard case of direct reference, the referent itself does not determine the epistemic content of the concept. In the phenomenal case, the referent plays a far more integral role:

⁸³ Note the use of “substantive” here when describing *R*, especially as it distinguishes *R* from *E*. Chalmers writes, “Phenomenal realists (e.g. Loar 1997; Hawthorne 2001) analyzing what Mary learns have occasionally suggested that her phenomenal concept is a demonstrative concept. This is particularly popular as a way of resisting anti-materialist arguments, as it is tempting to invoke the distinctive epistemic and referential behavior of demonstrative concepts in explaining why an epistemic gap does not reflect an ontological gap. But on a closer look it is clear that Mary’s central phenomenal concept *R* (the one that captures what she learns) is *distinct* from her central demonstrative concept *E*, as witnessed by the nontrivial identity $E = R$, and is not a demonstrative concept in the usual sense. This is not just a terminological point. Those who use these analyses to rebut anti-materialist arguments typically rely on analogies with the epistemic and referential behavior of ordinary (Kaplan-style) demonstratives. Insofar as these analyses rely on such analogies, they fail. Something similar applies to analyses that liken phenomenal concepts to indexical concepts (e.g. Ismael 1999, Perry 2001). If my analysis is correct, then pure phenomenal concepts (unlike demonstrative phenomenal concepts) are not indexical concepts at all.”

the quality of the experience plays a role in constituting the epistemic content of the concept and of the corresponding belief. One might say very loosely that in this case, the referent of the concept is somehow present inside the concept's sense, in a way much stronger than direct reference. (2003: 233)

a phenomenal quality can be thought of as filling a slot that is left open in the content of a direct phenomenal concept, and thus contributing its content. If there is no phenomenal quality to fill the slot, as in [the zombie's] case, the concept will have no content at all. (2003: 235)

Dualists are also constrained by their commitment to the *near Cartesian certainty* of our phenomenal *beliefs*, our beliefs of our own occurrent phenomenal experience. This higher degree of certainty than could be guaranteed by a "mere" reliable or causal connection between knower and known. As Chalmers writes,

The knowledge that a reliabilist theory grants us seems too weak to count as the kind of knowledge that we have of our conscious experience...if our beliefs about consciousness were justified only by a reliable connection, then we could not be *certain* that we are conscious... the only way to be sure here would be to have some further access to the other end of the connection; but that would be to say that we have some further basis to our knowledge of consciousness. (1996: 194)

If our phenomenal beliefs were justified by an ordinary causal connection between ourselves and the object of our belief, it would leave the zombies' phenomenal judgments as justified as our own, an unacceptable consequence for a property dualist theory of phenomenal knowledge. "A property dualist should argue... that the justification of our judgments about consciousness does not lie with the (causal) mechanisms by which those judgments are formed" (1996: 193, parenthetical original).

A subject bears a special relation to the phenomenal quality of her experience, a relation that is unique in many respects, one of which is that it holds *only* between a subject and her phenomenal properties, and to no other properties she might have.

This relation would seem to be a peculiarly intimate one, made possible by the fact that experiences lie at the heart of the mind rather than standing at a distance from it; and it seems to be a relation that carries the potential for conceptual and epistemic consequences. We might call this relation *acquaintance*. (2003: 248)

Acquaintance, Chalmers claims, is the “relation that makes possible the formation of pure phenomenal concepts... and makes a sort of lucid understanding [of phenomenal properties] possible” (2003: 248).

...even if we allow that zombies have beliefs, it is clear that a zombie cannot share a conscious being's phenomenal beliefs. The content of a conscious being's direct phenomenal beliefs is partly constituted by underlying phenomenal qualities. A zombie lacks those qualities, so it cannot have a phenomenal belief with the same content [nor a phenomenal concept with the same content, for the same reason]. (2003: 257)

3.4b The Limits of the Zombie’s “Phenomenal” Concepts and “Phenomenal” Beliefs, and The “Zombie Intentional Object” Thesis

Given Chalmers’s account of how we come to have beliefs about phenomenal properties, the special acquaintance relation we bear to the contents of our conscious experience, and the stipulation that a zombie “experience” has phenomenal *quality*, it does seem to follow that zombie-beliefs and zombie-concepts would have no phenomenal *content*. As I quoted him a few pages ago, Chalmers writes, “If there is no phenomenal quality to fill the slot [that is left open in the content of a direct phenomenal concept], as in [the zombie’s] case, the concept will have no content at all” (2003: 235). This is not strictly speaking true- all concepts have some non-phenomenal content- some cognitive or functional component- and zombie concepts will share the non-phenomenal content of their human-twin’s corresponding concept.

While zombie beliefs cannot have phenomenal content in the same manner that our “direct” phenomenal beliefs have phenomenal content, that is, by being “partially composed” of the phenomenal property that is its object, it is not obvious whether zombies may nevertheless have beliefs *about* phenomenal properties. To put this point in the form of a question: could a zombie belief, desire, or other propositional attitude take phenomenal properties, or phenomenal consciousness, or any form of “phenomenality”, as its objects? Can any of a zombie’s propositional attitudes have phenomenal objects?

3.4b1 Actual Epistemic Access to Phenomenal Properties

If we accept the property dualist’s claim that no type of epistemic access short of *acquaintance* could account for the special intimacy we have with the phenomenal qualities of our experience, we could formulate criteria for having actual epistemic access to phenomenal properties in this way:

AEAq: For a subject, S, to have actual epistemic access to a phenomenal property, Q, S must be acquainted with Q.

I think it is uncontroversial to assume that, according to the property dualist, one of the essential properties of human beings is *being conscious*, or at least having the capacity for conscious experience. If *being conscious* were not essential to *being a human*, it would be possible for a human being to be a zombie. And this is precisely what the property dualist wants to deny- according to property dualism, it is *conceptually necessary* that we, *qua human beings*, are conscious creatures. And since the zombie, by definition, is not phenomenally conscious, then the *absence* of phenomenal consciousness is an essential property of the zombie. By definition, for a zombie to be a *zombie* there must be “nothing it’s

like” to be him. If my zombie-twin were conscious she would not be my zombie. So assumption (1) below ought to hold:

Assumption 1. “Not being acquainted with phenomenal properties” is an essential property of the philosophical zombie.

3.4b2 Potential Epistemic Access to Phenomenal Properties

I believe it also makes sense to say that a subject S has *potential* epistemic access to Q only if it is in principle possible for S to have *actual* epistemic access to Q. And in these criteria there is an implicit but appropriate assumption that for S to perform some task Phi, S must be able to Phi without, for lack of better words, becoming ~S. Without this implicit assumption, the S who does Phi would not be S, because, in doing Phi, S would be ~S. And this should be conceptually incoherent. Regarding phenomenal properties, we would say that S has potential epistemic access to Q only if it is logically possible for S to have *actual* epistemic access to Q *while remaining S*, or *without becoming a non-S* or ~S. To put this implicit assumption in terms of having *potential epistemic access* to phenomenal properties,,

PEAq: For a subject, S, to have *potential epistemic access* to Q, it must be logically possible that

- (1) S has actual epistemic access to Q
- (2) In virtue of having epistemic access to Q, S is not ~S

It is clear where these criteria for “epistemic access” are leading. If a subject S has phenomenally conscious experience, S is not a zombie. And to have phenomenal conscious experience is to be acquainted with that experience, a kind of epistemic relation that a zombie does not bear to any of his zombie-states, and the grounds for Assumption (1). Since these are all stipulated criteria for having epistemic access to phenomenal properties, we can say

that, within this stipulated framework, it *conceptually necessary* that if S has actual epistemic access to phenomenal property Q, then S is *acquainted* with Q, and it follows that S is not a zombie.

This is where the criteria for potential epistemic access become important. The criteria for S having potential epistemic access to phenomenal properties required that it be logically possible for S to have actual epistemic access to phenomenal properties. So if S has potential epistemic access to phenomenal properties, S is not a zombie. Within this framework, it is conceptually necessary that zombies *do not have potential epistemic access to phenomenal properties*.

Our original question was whether zombies could have beliefs about phenomenal properties, or consciousness, or qualia, or whether the *object* of any zombie propositional attitude could have phenomenal qualities, given that the zombie's "phenomenal" belief itself cannot have phenomenal qualities. And if having a belief about Q requires having either actual or potential epistemic access to Q, then the answer is no. Zombies cannot have beliefs that take anything that has phenomenal content as their object, because they have no epistemic access to "the phenomenal". This would mean that no zombie-belief, neither a true zombie-belief nor false zombie-belief, is *about* the kind of objects that our corresponding phenomenal beliefs, both true and false, take as their object. The zombie cannot believe he is conscious, nor can he disbelieve he is conscious, nor be confused about consciousness, nor doubt that he is conscious. In short, zombies cannot think about "phenomenality" in any way.

More formally, I would like to put the point this way. Take "IO" below to mean "intentional object", the object of a subject's propositional attitude. And take z to qualify that

intentional object as one that belongs to a zombie. From the two principles above, AEAq and PEAq, we can form the “Zombie Intentional Object” thesis, which would look something like this:

3.4b3 The “Zombie Intentional Object” Thesis

IOz: It is conceptually incoherent for S to have both properties (1) and (2) below:

(1) is a zombie.

(2) has propositional attitude A with a phenomenal object Q.

3.5 The Invasion of the “Schmenomenal”-Zombies and The “Schmenomenal” Concept Strategy

If the zombie’s phenomenal concepts cannot have the same *content* as our own since, by hypothesis, they lack phenomenal states. This in itself does not prohibit the zombies from sharing our epistemic situation, as Chalmers notes,

the claim that a zombie and conscious being share their epistemic situation does not require that their beliefs have the same content. This mirrors the general requirement that epistemic situations be understood in topic-neutral terms. (2007: 177)

If our phenomenal concepts and beliefs and the zombies’ “phenomenal” concepts and beliefs differ in content, there are two general options for a physicalist account of zombie “phenomenal” concepts. (1) Zombie concepts are contentless- there is *nothing* taking the place of the phenomenal properties as the referents of the zombie’s analog concepts, or (2) that the zombies have “contentful” properties with some *non-phenomenal* content to which their “phenomenal” concepts refer.

Carruthers and Veillet (2007) reply to Chalmers’s “dilemma” by taking up the second option. The zombies’ “phenomenal” properties and concepts would not be identical to our

own, but these analogue non-phenomenal zombie-properties and the zombie-“phenomenal” concepts that refer to them could explain why, for example, zombies *say* there is an epistemic gap between the physical and phenomenal, and why they *say* they can conceive of zombies and inverts. If these concepts are sufficiently similar our own phenomenal concepts, and these “phenomenal” beliefs have the same epistemic status as our phenomenal beliefs, then we should conceive of our zombie-twins as sharing our epistemic situation.

Carruthers and Veillet consider, and reject, the interpretation of zombie-phenomenal concepts on which these concepts refer to *our* phenomenal states; if that were so, any claim a zombie might make about her own phenomenal state would be false. They claim it is more appropriate to equate the zombies’ “phenomenal” concepts to Oscar’s and Twin Oscar’s concept(s) of ‘(t)water’, a paradigm case of two subjects “sharing an epistemic situation” though differing in the content of their beliefs. Oscar’s and Twin-Oscar’s corresponding ‘water’ beliefs have the same truth value and same epistemic status- they are both cognitively significant and they are equally justified- thereby sharing the same epistemic situation. As Carruthers notes, if we were to insist that “water” as used on Twin Earth referred to H₂O because ‘water’ refers to H₂O *as we use it*, Twin Oscar would speak falsely every time he mentions *water*.⁸⁴ Twin Oscar would be “...wrong every time he says, ‘This water tastes good.’ But clearly that is just absurd. No theory of concepts does (or should) yield such a counterintuitive claim” (2007: 17). If no theory of concepts should demand that Twin Oscar’s *water* concept to refer to H₂O, to Oscar’s version of water, the, according to

⁸⁴ Twin Oscar would also have no epistemic access to his error, something that will become important in the later parts of this discussion.

Carruthers and Veillet, no theory of phenomenal concepts should allow that a zombie's "phenomenal" concepts refer to our (version of) phenomenal states.

we don't need to claim that my zombie-twin and I deploy the same... concept in order for our points to go through. It just has to be the case that the zombie deploys a concept that is like mine in the relevant functional-role respects. For it turns out that it is this (conceptually isolated functional) role that is sufficient to explain the conceivability of zombies, the appearance of an explanatory gap, and so forth. (ibid: §2.3)

In any discussion of zombies, it is agreed on all sides that zombies are at least *functional* duplicates of human beings. So it ought to be just as readily conceded, for any human concept, zombies have features that functionally correspond to concepts, and whatever functional roles our concepts might play in our cognitive economy, zombies have concepts that play the relevant, functionally equivalent, roles in their cognitive economy. Assuming that phenomenal concepts, though conceptually isolated from physical and functional concepts, do play some sort of role in our cognitive economy (the "phenomenal cognitive economy", perhaps),⁸⁵ then whatever role that may be, zombies will have a concept that plays the same role in their cognitive economy as our phenomenal concept plays in ours.

If the physicalist could show that the zombie, properly conceived, has a functional analogue of our phenomenal concept, and these analogue concepts allow the zombie to have "phenomenal" beliefs with the same truth value and the same epistemic status as our own, then Carruthers's and Veillet's approach could disarm the second horn of Chalmers's dilemma (that P&~C is *not* conceivable). If we do conceive of zombies as sharing our

⁸⁵ Phenomenal concepts may have functional roles insofar as, e.g., one can make inferences based on beliefs involving phenomenal concepts to other beliefs involving phenomenal concepts. Of course we are assuming that this is an isolated area of our cognitive economy, so that one cannot make inferences from beliefs couched in phenomenal terms to beliefs involving only non-phenomenal concepts.

epistemic situation, as Carruthers and Veillet note, “at least in one good sense of the notion of ‘epistemic situation’” (ibid, §4), the physicalist need not explain why we can conceive of physical duplicates of ourselves who are in a different epistemic situation from ourselves.

In possessing these contentful, functional-analogues of our phenomenal concepts, Carruthers and Veillet claim that zombies and humans can thereby share their *epistemic* situation (though not their *phenomenal* situation), and the zombie-counterparts of our phenomenal concepts will play the same role in explaining *their* epistemic situation (e.g., why zombies *say* there is an explanatory gap in their world) as our phenomenal concepts play in our own case. If we do share our epistemic situation with our zombie-twins, explaining ours and theirs with the same “tool” will not be a flaw in the phenomenal concept strategy.

Chalmers is willing to grant for the sake of argument that there may be a conceptually coherent zombie-type creature who has these ersatz “phenomenal” properties- “schmenomenal properties”- and that their “phenomenal” concepts could be “schmenomenal” concepts that refer to schmenomenal states. If zombies did have such a state, then,

just as Mary gains cognitively significant non-indexical knowledge involving phenomenal concepts, Zombie Mary gains analogous cognitively significant non-indexical knowledge involving schmenomenal-concepts.... Zombie Mary gains significant knowledge of the form *Tomatoes cause such-and-such schmenomenal state, I am in such-and-such schmenomenal state, and This state is such-and-such schmenomenal state*. Zombie Mary’s new beliefs have the same truth-value, the same epistemic status, and the same epistemic connections as Mary’s corresponding beliefs. (Chalmers, 2007: 186)

But Chalmers denies that the conceptual coherence of this kind of creature implies that we share our epistemic situation with zombies, and claims this interpretation of zombie-knowledge (with schmenomenal-concepts composing schmenomenal beliefs about

schmenomenal states) is an inappropriate interpretation- an “inflation”- of the knowledge a philosophical zombie (properly conceived) could possess.

when we ordinarily conceive of zombies, we are not conceiving of beings with something analogous to consciousness that is epistemically just as good. Rather, we are conceiving of beings with nothing *epistemically analogous* to consciousness at all. (2007: 186, emphasis mine)

“Schmenomenal-zombies” are not the property dualist’s zombie: according to the property dualist, where we have phenomenal properties the zombies have *nothing*- no property or state epistemically equivalent to our phenomenal properties and states. However, it should be noted that the epistemic insignificance (or cognitive insignificance) of zombie beliefs- whatever their content (or lack thereof)- does not follow *directly* from the original definition of a zombie; if it does, that entailment certainly is not obvious. Our epistemic inequality may well be a dualist addendum to the otherwise agreed upon definition of a philosophical zombie, one that the physicalist should not accept without argument.

If our zombie counterparts do share our same epistemic situation, then when a zombie acquires the zombie analogue of a phenomenal concept, he would acquire the zombie analogue of our phenomenal *knowledge*, which would be as substantial or insubstantial, cognitively significant or insignificant, as our phenomenal knowledge. While you have knowledge of your phenomenal states, your zombie-twin has knowledge of his *schmenomenal states*: “states that have the same physical, functional, and intentional properties as Chalmers’s [phenomenal] states, but that aren’t phenomenally conscious” (2007: 9). Where we have beliefs about *phenomenal* consciousness, “zombies are thinking about schmenomenal consciousness using their phenomenal [schmenomenal] concepts, which are conceptually isolated from their other [non-schmenomenal] concepts” (ibid: 10).

The conceptual isolation of schmenomenal-concepts from physical concepts explains why zombies believe that there is an explanatory gap between physical and phenomenal facts. On this view, there really *is* an explanatory gap in the zombie world, and it exists for the same reason that the explanatory gap exists in our world: just as our physical concepts and phenomenal concepts are isolated, the zombie's physical and schmenomenal-concepts are isolated, creating the same epistemic effects.

3.5a The Zombies' "Schmenomenal" Beliefs

Whatever zombie-"phenomenal" statements actually *mean*, they have no phenomenal content and do not refer to phenomenal properties. Zombie-phenomenal concepts are not *phenomenal* concepts at all; rather, they should be understood as "schmenomenal-concepts", referring to "schmenomenal" properties. On this interpretation of zombie-phenomenal talk, "schmenomenal" statements share the truth-value of their corresponding phenomenal statements. Whenever a zombie's human twin speaks truly, the zombie speaks truly; when her human twin statement is false, the zombie's statement is false.

This is a popular a posteriori physicalists interpretation of the content of zombie-phenomenal concepts. Zombie-"phenomenal" beliefs are not phenomenal but on this interpretation *schmenomenal*; schmenomenal beliefs may be true, but their relation to phenomenal belief, as we have characterized it, is complex. When your zombie-twin says, "I am phenomenally conscious", or "Now I know what it's like to see red!" he says something true, but his statement has different content and the intended referent is not what yours would be if you were to utter those same phonemes. Your zombie-twin does not mean what you would mean if you were to say "Now I know what it's like to see red!" When a zombie utters,

"I am experiencing phenomenal-redness", she may be saying something both meaningful and true in virtue of successfully referring to a schmenomenal property she is currently instantiating. If the zombie-phenomenal concept refers to a zombie-mental property, (functionally identical to the referent of a phenomenal concept) then the zombie-phenomenal-assertion may be true or false, depending on the nature of this referent and whether the reference is successful. As Carruthers and Veillet write,

Zombie Chalmers is correct when he says that he is conscious, because he isn't saying that he has phenomenal states as we understand them. He is correct because he means that he has schmenomenal states, and he has them. (2007: 20)

This option is attractive for many reasons. Perhaps the most "psychologically" compelling is that it makes sense of zombie life, of *why* zombies say what we say and behave just as we behave in corresponding situations. It makes the zombies' behavior, zombie-talk, *intelligible*, rather than mind boggling. When my zombie-twin says "I am experiencing phenomenal-redness," she means something *other* than what I mean, and she is correct. When Zombie Mary leaves the black and white room, she *does* learn something, and that's *why* she's surprised- but what she learns is not identical to what Mary learns when she leaves her room. Zombie-Mary's new "experience", her new knowledge, does not have phenomenal content; her new state lacks *phenomenal feel*.

3.5a1 Schmenomenal Beliefs- True, But Complex

Zombie-"phenomenal" concepts differ from human phenomenal concepts in that they neither refer nor purport to refer to phenomenal properties, the referents of our phenomenal concepts.

Zombie-"phenomenal" concepts refer to "schmenomenal" properties (properties that exist in the zombie world and have the physical, functional, and intentional features of human phenomenal properties), *and the truth value of a zombie's "schmenomenal" assertion will correspond with her human twin's phenomenal assertions.*

On the “schmenomenal” interpretation, zombies have neither phenomenal concepts nor phenomenal beliefs, nor do their concepts or beliefs *purport* to have such content. This makes some intuitive sense; it is arguable that the existence of phenomenal properties in *our* world should have no bearing on the content of any concept or belief in the *zombie*-world, since phenomenal properties are wholly absent from the zombie-world, and the property dualist claims that the zombie has no property with which to replace them (no property they do not share with their human twins). As always, there will be some molecule-for-molecule, phonetically identical, zombie utterance corresponding to every human assertion, but the content of the zombie and human expressions are determined independently of one another and depend solely upon facts about their own world.⁸⁶

3.6 Epistemic Equality for All?

The phrases “epistemically identical” and “epistemically equivalent” may be used differently in different contexts; as I am using the term here, beliefs are epistemically *identical* when they have the same or appropriately corresponding truth conditions, when they have equal strength of justification and appropriately corresponding sources of justification. Oscar’s and Twin Oscar’s beliefs about “water” and “(t)water” are a paradigm case: Oscar’s water-beliefs and Twin Oscar’s twater-beliefs have the same truth value and

⁸⁶ In order to give the zombie-phenomenal concept independent content, content determined independently of the content of our own phenomenal concepts, we must treat the zombie-world as actual, as a way the world could have been.

One might object that it is inappropriate to treat the zombie-world as anything but a counterfactual context of evaluation for our own phenomenal concepts, and that the only appropriate world to treat as “actual” is one in which “phenomenal” concepts refer to *qualia*, to the properties that constitute the character of experiences that have a *subjective feel*. We will consider a similar objection later (though *prima facie* it seems that this objection assumes more than the anti-materialist is entitled to assume).

corresponding truth conditions, the justification for each of the twins' beliefs is equally strong, and, depending upon the degree of "twin-ship" between Earth and Twin Earth, they will have functionally and epistemically identical sources of justification as well⁸⁷. Their beliefs are epistemically identical though they may differ in content. Oscar and Twin Oscar might be so similar as to share their entire epistemic situation, making them epistemic twins.

3.6a Schmenomenal-Zombies: Our Epistemic Twins or Epistemic Equals?

In the spirit of the two Oscars, we can call subjects whose cumulative sets of beliefs are epistemically identical "epistemic twins"; if it is possible for *any* two subjects share their entire epistemic situation, *epistemic twins* certainly will. But the criteria for epistemic identity is, by definition, extremely demanding, and it seems overly demanding that two subjects' beliefs must be epistemically identical in order for those subjects to share more circumscribed aspects of their overall epistemic situations.

There should be an additional category, call it "epistemic equality", for beliefs that have corresponding but distinct "correctness conditions", whether these are truth conditions, or appropriateness conditions, or fitting another kind of norm. And since there may be multiple sources of justification for different beliefs, all of which could lend the same strength of justification, beliefs which can arguably be said to share their epistemic status

⁸⁷ The source of justification for Oscar's water belief is his connection to H₂O; the source of Twin Oscar's (t)water belief is his connection to XYZ. If these connections are functionally identical, their sources of justification are functionally and epistemically identical.

even though their *sources* of justification may not be shared should be “epistemically equal” in this aspect as well.

Corresponding schmenomenal and phenomenal beliefs might be epistemically identical, epistemically equivalent, or neither; so far this question has been left open. They might differ in their truth conditions, or one belief may be justified when the other is unjustified or less justified, or there may be some other important epistemic difference between them. In Chalmers’s “dilemma” for the phenomenal concept strategy, he argues for the epistemic *inequality* of phenomenal and schmenomenal beliefs; schmenomenal beliefs are either false while our corresponding phenomenal beliefs are true, or less justified, or in some way epistemically inferior. Whatever the relevant epistemic difference may be, the anti-materialist is committed to the position that he does not share his epistemic situation with his zombie-twin. But if schmenomenal-concepts refer to schmenomenal properties, properties present in the zombie world, it seems that there is logical space for the epistemic equivalence of schmenomenal and phenomenal beliefs such that you and your zombie-twin could be epistemic equals.

So far we have one concrete example of epistemic-twinship: the Oscars. Any purported example of epistemic equality *without* epistemic identity will be more contentious. By my definition of epistemic equality above, subjects may hold epistemically equivalent beliefs so long as those beliefs have equal epistemic *status*, though the source of justification might differ substantially. Just *how radical* the difference between the sources may be before that difference entails different strengths of justification on corresponding sets of beliefs is an

open question. I believe the following would count as a non-radical difference between sources of justification:

You and I read two equally well regarded textbooks, A and B, on cognitive neuroscience, covering sufficiently similar material, we may have epistemically equivalent beliefs about the content of the material covered. If you attend a lecture given by the author(s) of A, discussing the same material as before, while I read the corresponding material from textbook B, you and I may still hold epistemically equivalent beliefs about the material discussed.

These differences seem unproblematic, fairly insignificant. Examples of more radical differences in sources of justification that still result in epistemically equivalent beliefs are much more difficult to argue. Rather than making a straightforward *positive* case for subjects' holding epistemically equivalent beliefs with radically different sources of justification and/or corresponding but distinct "correctness conditions", it will be easier to reconstruct- in order to deconstruct- particular arguments for the epistemic *inequality* of phenomenal and schmenomenal beliefs.

The success of the conceivability argument, the knowledge argument- of *any* anti-materialist argument grounded in the "explanatory gap"- depends on whether our epistemic situation differs from that of the zombie. If there is no compelling argument against the epistemic equality of phenomenal and schmenomenal beliefs, the dualist will be in a serious bind. Without such an argument, he cannot deny the physicalist's claim that humans and their schmenomenal-zombie-twins share their epistemic situation. Undermining arguments against the epistemic equality of human and zombie beliefs may be the most compelling way to show that we share our epistemic situation with our zombie-twins. If the physicalist can do this, she will have shown that there is a fully physical account of our epistemic situation, thus showing that *our* epistemic situation provides no evidence against materialism; that is, the

“explanatory gap” gives us no reason to believe that property dualism is true of *our* world while false in the zombie’s.

3.6a.1 The Marys and The Oscars

Consider again the case of Oscar and Twin Oscar. When the Oscar and Twin-Oscar learn “water is wet,” they arguably acquire epistemically equivalent beliefs. Their beliefs have corresponding truth-value, seem equally justified, and both are cognitively significant. Oscar and Twin Oscar satisfy the dualist’s criteria for sharing an epistemic situation though their beliefs are about *different objects* and involve acquiring *different concepts*⁸⁸.

Is the Mary/Zombie Mary scenario epistemically analogous to that of the Oscars’ ‘water’ and ‘(t)water’ beliefs? Is the phenomenal knowledge Zombie Mary acquires when she sees color for the first time epistemically equivalent to Mary’s new phenomenal knowledge? Carruthers and Veillet believe so, since it is not required that corresponding human and zombie beliefs share their *content* in order for the twins to share their epistemic situation; epistemic situations are to be thought of in “topic neutral” terms.

Can the property dualist can give a compelling argument for the view that humans and phenomenal-zombies would not share their epistemic situations, an argument consistent with the dualist’s commitment to topic-neutrality here and the definition of a zombie? In the following section we will look at three possible arguments: the first appeals to a difference in the abilities of phenomenally conscious beings and their zombie-twins; the second to the *richness* of phenomenally conscious experiences unrivaled by any zombie-

⁸⁸ This is true at least on a “wide content” interpretations concepts: Oscar acquires the concept, *water*, and Twin Oscar acquires a distinct concept, *(t)water*.

state; and the third examines a possible difference between the *strength of justification* of phenomenal beliefs and the beliefs a zombie might possess.

3.7 Three Property Dualist Arguments Against Epistemic Equality

3.7a An Illicit Appeal to Functional Disparities

Chalmers claims that only if one assumes that some naturalistic theory of mind is true *a priori* can the physicalist argue that humans and zombies are in the same epistemic situation.⁸⁹ He writes,

Zombies have the same functional organization as conscious beings and the same reliable causal connections among their physical states, so a proponent could suggest that these theories entail that zombies will have corresponding beliefs with the same epistemic status as ours... But in any case, to appeal to these theories in this context is to beg the question. Considerations of the Mary situation and related matters gives us good reason to believe that consciousness is relevant to matters such as mental content and epistemic status. (2007: 187ff)

Not necessarily. Considerations of the Mary situation and related matters do not give us *unequivocally good* reasons to think consciousness is relevant to the comparative epistemic status of our phenomenal beliefs and the zombie's phenomenal beliefs. Both sides agree that consciousness is relevant to mental *content*—in conceiving of a zombie, one conceives of a creature with *no phenomenal consciousness*, whose mental content is *different* from that of a phenomenally conscious being. So the physicalist can accept that phenomenal consciousness is relevant to mental content without *thereby* conceding that this distinction

⁸⁹ These would be theories such as a functionalist theory of belief, a causal theory of mental content and/or a reliabilist theory of knowledge.

entails a difference between Mary's and Zombie Mary's epistemic situations. It has already been established that a difference in the content of subjects' beliefs does not entail a difference in the epistemic status of those beliefs or a difference in the subjects' epistemic situations.

Property dualists and a posteriori physicalists agree that *Zombie Mary* gains some kind of new information upon seeing red for the first time, and that this event and resulting zombie-knowledge is functionally equivalent to whatever Mary learns when she first sees red. But if we take Chalmers's (2010) interpretation of Mary's new knowledge here as an explanation of their epistemic difference, we should throw out considerations of the "Mary situation" entirely. He writes,

As for *Zombie Mary*'s "new knowledge," it is clear that she gains no propositional knowledge (though she may think that she does)... When *Zombie Mary* first sees a flower, she may gain certain abilities to recognize and discriminate, *although even these abilities will be severely constrained since they cannot involve experiences.* (Chalmers, 2010: 295, emphasis mine)

By hypothesis, none of *Zombie Mary*'s properties involve conscious experience, but nothing about this passage illuminates *why* acquiring knowledge involving conscious experience and acquiring functionally identical knowledge without phenomenally conscious experience makes a difference in the *epistemic status* of Mary's and *Zombie Mary*'s "new knowledge".⁹⁰ What this passage *does* illuminate is one of the unsuccessful ways in which the dualist might attempt to argue for a difference between Mary's and *Zombie Mary*'s

⁹⁰ We may or may not want to grant Chalmers's first claim here, that *Zombie Mary* gains no new propositional knowledge; by hypothesis, *Zombie Mary* will believe* she gains propositional knowledge, just as Mary does, and whether or not this belief* amounts to actually gaining propositional knowledge is an open question. Note: A belief* is functionally equivalent to a belief, but stipulated as belonging to a creature who has no phenomenal properties, and the belief* itself stipulated as having no phenomenal content.

epistemic situations, that is, by claiming there will be a difference in the *abilities* they gain upon perceiving color for the first time. Positing a discrepancy between these twins' abilities would violate the definition of the relation between humans and their zombie-twins; dualists and physicalists alike should reject such a claim.

Chalmers writes that Zombie Mary's new abilities to recognize and discriminate colors will be severely constrained because they do not involve color experiences. If Zombie Mary's abilities *were* so constrained, we would have a compelling reason to believe that Zombie Mary's "new knowledge" was epistemically inferior to that of her human twin. However, Zombie Mary's abilities could only be constrained *relative to* another epistemic agent's abilities, and by hypothesis Zombie Mary's abilities to perform such cognitive functions as *recognizing* and *discriminating* cannot be constrained relative to Mary's new abilities. By hypothesis, the ability to discriminate, recognize, retain information, recall information, make identifications, are all *physically reducible* mental properties, and according to the property dualist's definition of a philosophical zombie, your zombie-twin's physically reducible mental properties are identical to your own. This will include physically reducible mental properties, including your physically reducible cognitive ability to recognize and discriminate colors. Zombie Mary's ability to recognize and discriminate colors cannot be "severely constrained" as compared to her human twin; on the contrary, such abilities are necessarily identical between human-zombie-twins. Any other

interpretation of Mary's and Zombie Mary's corresponding abilities would violate the dualist's own stipulated definition of a philosophical zombie.⁹¹

Whatever difference there may be between what happens to Mary and to Zombie Mary when they first see color, this difference cannot give rise to a difference between their cognitive abilities or dispositions regarding colors. Their new knowledge must be psychologically identical in Chalmers's functional sense of "psychological". This knowledge plays the same role in their cognitive economy, thus, as Carruthers and Veillet argue, this aspect of the Mary-scenario would be evidence in *favor* of Zombie Mary and Mary having epistemically equivalent beliefs, or sharing their epistemic situation, "at least in one good sense of 'epistemic situation'".

3.7b "This" is Rich- A Second, Less Obvious, Illicit Appeal to Functional Disparities

Rather than appealing directly to functional disparities between our phenomenal knowledge and a zombie's schmenomenal knowledge, perhaps the dualist can ground the difference in our epistemic situations in the difference between the *richness* of our respective inner lives. Chalmers claims that,

...when we conceive of zombies, we are not conceiving of beings whose inner life is as *rich* as ours, but different in character. We are conceiving of beings whose inner life is dramatically *poorer* than our own... this difference makes for a dramatic difference *in the richness of our introspective knowledge*. (1996: 186, emphasis mine)

⁹¹ This is the kind of disquieting misinterpretation of the relation between human beings and their zombie-twins that leads some philosophers, notably, Daniel Dennett, to clam that such a thought-experiment "actually encourage us to misunderstand its premises!" (1991: 398).

The physicalist who employs the phenomenal concept strategy grants that human lives are full of rich phenomenal content while the zombie's life, by definition, has none.⁹² Of course, the fact that our inner lives are richer than those of zombies does not straightforwardly imply that this "richness" makes an *epistemic* difference regarding our knowledge of our inner states. To make this claim, there must be some further argument.

Chalmers has elsewhere appealed to the "*richness*" of our phenomenal states and the "poverty" of the content of the zombie's schmenomenal states to illustrate a difference between the two kinds of mental states in a way that *suggests* an *epistemic* difference between *rich* phenomenal beliefs and "poorer" schmenomenal beliefs. The adjective "richness" often appears in discussions of the qualities of our introspectable mental states and those of the zombies, and here it is at the very least it is *implied* that there is an epistemic difference between these two kinds of belief which follows from the richness of phenomenal experience compared to whatever "introspective" knowledge a zombie might have.

I believe it would be a mistake for the property dualist to use this purported difference in the "richness" of our mental lives to ground an epistemic difference between corresponding phenomenal and schmenomenal beliefs. All parties have accepted that we conceive of zombies as "phenomenally barren"; their introspectable states lack any phenomenal content whatsoever. Where we have phenomenal states, they have *nothing*. But it would be a mistake to think of schmenomenal states as "impoverished phenomenal states". What sort of poverty would this be? By definition schmenomenal states bear *no qualitative*

⁹² It is worth noting that in acknowledging that though we can conceive of ourselves as having conscious experience, phenomenal states, or qualia, nowhere in the zombie hypothesis is there a condition that states we must conceive of these states as "rich" in some epistemically significant way.

resemblance to phenomenal states, and unlike any other form of poverty, zombies cannot *suffer* from their “phenomenal poverty”. Not only do they have no *idea* what they’re missing; by definition, if a zombie suddenly inherited a fortune in qualia, she wouldn’t -she couldn’t- even notice.

A sunset might have rich, saturated colors; a béarnaise sauce might be “rich”, full of oil and egg. This is the sense of “rich” that Chalmers emphasizes when he describes conscious experience as “rich”, yet there is a way of using “rich”- perhaps it is homophonic, perhaps metaphorical- to mean “wealthy”, possessing more than adequate resources for a given task.⁹³ Though this dual sense of “richness” may seem insignificant, I believe that the argument from the *richness* of phenomenal knowledge to the *epistemic inferiority* of “schmenomenal” knowledge in fact turns on an equivocation between the sensorial and functional senses of “rich”.. No successful argument can be based on an equivocation between two meanings of a term. I believe an argument from “richness” would be just such an argument, beginning with a claim about the sensorial “richness” of phenomenal states as compared to schmenomenal states, and infer from this a non-sensorial, epistemic difference.

Consider the connection between *richness* and *wealth*. We sometimes describe an individual as having a “wealth” of knowledge, or one’s knowledge of a subject as “richer” than another’s. But I would contend that this sense of “richness” and “wealth” is implicitly functional. When I say that Jon has a “richer” knowledge of French literature than I do, I believe it is most plausible that my statement is elliptical for something like, “Jon has *more* and *better justified* knowledge” of French literature than I do- perhaps more knowledge of

⁹³ Usually, this task is merely “living the good life”.

certain details, of the significance and influence particular works, of events surrounding them, etc. For Jon's knowledge of French literature to be epistemically richer than mine would be a *functional* difference between our respective French-literature-beliefs, and there can be no *functional* difference between corresponding human and zombie beliefs, even though there may be a difference in their *content*.

Monetarily, "wealth" is something like a collection of assets- properties and objects, stocks, the contents of a Swiss bank account- and it might seem that "wealth" in this sense is *not* functionally determined. But there is still an implicit connection between that which makes a person wealthy, monetarily, and the functional definition of the term "rich". A person can lose their *wealth*- their properties, objects, and stocks, can lose *value*- in an economic downturn, though metaphysically, and even physically, the property and material objects that constitute that wealth have not themselves changed. If you once owned 1000 shares of Google at 500 points per share, and held them through a stock market decline, you own "the same" 1000 shares post-downturn, but your *wealth* is lost; your assets have become less valuable because their economic worth has decreased, and economic worth is determined functionally.⁹⁴

⁹⁴ There are plenty of (I believe harmless) *prima facie* counter examples to a functional definition of "rich". Is a billionaire whose assets have been frozen still as *rich* as she was before the freeze? If the assets are only temporarily frozen and then released, was the woman rich in the beginning, then (during the freeze) did she stop being rich, and become rich again once she had access to her funds? Is a multi-millionaire in a permanent coma still rich? (We should assume that his funds are not paying for his care or in some other way being dispersed according to his wishes; for example, they might be used to further some project or desire he had before falling into the coma.) I don't believe that these questions have a single right answer; either "yes" or "no" might be correct depending upon the context and what feature of the situation one wants to emphasize. Out of spite one might say something like, "Oh, she's still a rich woman alright. She just can't touch a cent of that money!" The woman might be "rich" in some sense but clearly not in the important sense of "rich", the sense in which one can *do something* with their assets.

If phenomenal knowledge could be “rich” in the functional sense, it would be more apparent how this *richness* and *poverty* could create an epistemic difference between phenomenal and schmenomenal beliefs. One could *do* more, epistemically speaking, with phenomenal knowledge; one could make more or better inferences, for example, than one could make by acquiring schmenomenal knowledge.⁹⁵ But the “richness” of phenomenal consciousness, the richness of the qualitative character of an experience, is an “intrinsic” property of that experience- it is not an extrinsic *functional* property of the experience- and as we established above, phenomenal states cannot be *richer* than schmenomenal states in the functional sense.

Perhaps we should not read Chalmers’s quote above as making the claim that the difference in the richness of human and zombie inner mental states makes a *direct* epistemic difference between phenomenal and schmenomenal knowledge. Often talk of “richness” transitions into talk of “substantiality”, for example, to continue the quote above,

Where we have substantial knowledge of our phenomenal inner lives, zombies have no analogous introspective knowledge: there is nothing analogous for them to have introspective knowledge of. (ibid: 186)

Perhaps our phenomenal knowledge is “substantial” *because* our inner lives are “rich”. Even here, Chalmers seems to make a questionable move from a claim about phenomenality to epistemology, one that the physicalist might resist. The passage begins with a description of the kind of knowledge that we have, not rich or robust or florid here but “substantial”. We have *substantial* knowledge of our phenomenal inner lives, and zombies

⁹⁵ By hypothesis, one cannot make inferences about the phenomenal quality of another individual’s experience without phenomenal knowledge, but I doubt this is the epistemic difference between phenomenal and schmenomenal states that Chalmers intends.

have no analogous introspective knowledge. Certainly-- by hypothesis-- zombies have no phenomenal knowledge. But Chalmers moves from this claim about the substantiality of our phenomenal knowledge to the assumption that the zombies have no inner *non*-phenomenal states of which they can have analogous, *substantial*, non-phenomenal, introspective knowledge.

It would be easier to see whether this move from “no substantial knowledge” to “no analogous inner introspectable state” was legitimate if we knew how to read “substantial” here. It is often used interchangeably with “important”, “cognitively significant”, “non-trivial”, or “non-indexical”, such as in the claim, “Mary gains substantive knowledge when she has an experience of color”. If we should interpret “substantiality” here in any of these ways, then the zombies’ schmenomenal knowledge is certainly “substantial”- their schmenomenal knowledge is as “important” to the zombies, as non-trivial and as *significant* in their lives as our phenomenal knowledge is in ours.

3.7c Does Our “Acquaintance” with our Qualia Justify (establish, create) Epistemic Inequality?

There is third possible source of epistemic inequality which we have not yet addressed: the *justificatory status* of a belief. By this I mean there may be a difference between the strength of justification for corresponding phenomenal and schmenomenal beliefs. It is possible that our phenomenal beliefs are better justified than our zombie-twins’ corresponding schmenomenal beliefs. And given the definition of “epistemic situation” with which we have been working, if our corresponding phenomenal and schmenomenal beliefs differ in their strength of justification, then we do not share our epistemic situation with our

zombie-twins, even if the *truth value* of their schmenomenal beliefs corresponds perfectly with the phenomenal beliefs of their human twins..

Chalmers alludes to such a strategy when he writes,

If the account [of phenomenal concepts and phenomenal belief] applies equally to a zombie... then it cannot account for the crucial epistemic differences between conscious beings and zombies. And if it does not apply equally to a zombie (if it relies on a notion of acquaintance, for example), then crucial explanatory elements in the account will not be physically explicable. (2007: 183, text in brackets mine)

“Acquaintance” is the special epistemic relation that holds between conscious subjects and the phenomenal properties instantiated in their experience- a relation that subjects do not bear toward any of the other properties they might instantiate. In being “acquainted” with our own phenomenal experiences, our phenomenal beliefs can be justified “with something approaching Cartesian certainty” (ibid: 183).

An argument from “acquaintance” might be the most plausible argument for the epistemic inequality of phenomenal and schmenomenal beliefs. If the “crucial epistemic difference” between our epistemic situation and the zombies’ is the difference between the strength of justification for a subject’s phenomenal belief and how strongly a zombie’s schmenomenal belief can be justified, then the dualist may be correct, and there may be no satisfactory physicalist account of our epistemic situation.

On a physicalist account of phenomenal concepts and phenomenal belief, both phenomenal and schmenomenal concepts and beliefs would be formed by the same, equally reliable, causal or functional mechanisms. As such, the justification of corresponding human and zombie beliefs would be equally strong. No physicalist account of our epistemic situation could substantiate a disparity between the epistemic status of our beliefs about our

introspectively accessible phenomenal states and the epistemic status of a zombie's beliefs about her introspectively accessible schmenomenal states. And if our *acquaintance* with the content of our phenomenal experience *creates* the epistemic difference between our phenomenal beliefs and the zombie's schmenomenal beliefs, then *so long as the acquaintance relation is not physically explicable*, there can be no adequate physicalist account of our epistemic situation.

The physicalist may accept that we are "acquainted" with the contents of our conscious experience, and that this epistemic relation holds only between a conscious subject and the properties of her phenomenal experience without *thereby* admitting to the presence of non-physical "phenomenal" properties in our world. Conceptually, a relation that holds only between a conscious subject and the content of her experience would not "apply equally to a zombie" (2007: 183). By the definitions of the acquaintance relation and of zombies, this epistemic relation cannot apply to a zombie, since zombies have no phenomenal experience in which to instantiate properties with which they could be acquainted. So far this is unproblematic: we conceive of zombies as lacking phenomenal states, and it seems we can conceive of an epistemic relation- call it "acquaintance"- that is more intimate than any other epistemic relation. It also seems we can conceive of this relation holding only between a subject and her phenomenal states. None of this seems to entail the *possibility* of non-physical phenomenal states; as always, these points of agreement stop at the *conceivability* of non-physical phenomenal states, it does not proceed to their possibility, and certainly not to their logical necessity.

But Chalmers means something much stronger when he claims that this epistemic relation “does not apply equally to a zombie”. To highlight the questionable element of the quote above, I will repeat it here.

if [the account of our epistemic situation] does not apply equally to a zombie (if it relies on a notion of acquaintance, for example), then crucial explanatory elements in the account *will not be physically explicable*. (2007: 183, emphasis mine)

The physicalist should become suspicious at this point, and ask whether- under this description of acquaintance- it would be *logically possible* to be “acquainted” with a *physically-reducible or supervenient* qualitative state, or if one can only be “acquainted” with *non-physical* phenomenal states. If it were conceptually necessary for this acquaintance relation to hold only between subjects and *non-physical* states, then (1) any account of the epistemology of phenomenal belief that appealed to acquaintance would be one that the physicalist must reject and (2) the acquaintance relation *itself* could not do any work in anti-materialist arguments since it *assumes* that phenomenal properties are non-physical.

For a relation to be physically *explicable*, it is not necessary for all its *actual* relata to be physical; a relation will be physically explicable if it is merely *conceivable* for all its relata to be physical. So the claim that the acquaintance relation- a crucial explanatory element in an account of our epistemic situation- is *not physically explicable* is quite strong. For a relation to *resist* physical explanation, it must be *conceptually necessary* that at least one of the relata be *non-physical*. So it seems that “acquaintance”, the epistemic relation that creates the disparity between phenomenal and schmenomenal belief, the relation that “does not apply equally to a zombie”, the relation to which the physicalist supposedly cannot appeal, requires that a subject’s experience instantiate some *non-physical* properties with which she can be

“acquainted”. And the physicalist has no reason to allow this as a *premise* in an argument for the existence of non-physical phenomenal properties.

3.8 The Epistemic Equality of Zombies- Living to Fight Another Day

Whatever difference there might be between our epistemic situation and that of our zombie-twins, we know that it cannot be a functional difference. And I hope to have shown that the “richness” of our phenomenal knowledge does not *itself* entail a difference in the *epistemic status* of phenomenal beliefs and schmenomenal beliefs, nor does it entail a difference between the truth value of these corresponding beliefs. It seems that it is our “acquaintance” with our phenomenal states that creates the epistemic difference between our phenomenal beliefs and our twins’ schmenomenal-beliefs- a difference in the strength of their justification- and that this difference in justificatory strength is the last epistemic difference between our situation and that of the schmenomenal-zombie left standing. And as we noted above, the acquaintance relation itself cannot do any *work* in an anti-materialist argument, since least one of its relata is necessarily non-physical.

Both physicalists and dualists originally agreed that the *content* of our beliefs is irrelevant to the question of the comparative epistemic status of our beliefs and to whether we can share our epistemic situation with the zombies. As Chalmers states in the quote from above,

the claim that a zombie and conscious being share their epistemic situation does not require that their beliefs have the same content. This mirrors the general requirement that epistemic situations be understood in topic-neutral terms. (2007: 177).

To suggest otherwise would be to beg the question against the physicalist. But the dualist cannot keep his word here: there is no difference between our epistemic situations that can be drawn from considerations of our functional properties or the dissimilarity of our mental content itself. Given Chalmers's criteria for sharing an epistemic situation, it seems the only explanation available is one that appeals to the justificatory status of our corresponding beliefs, and to establish a justificatory disparity we must assume that the relation between a conscious subject and her phenomenal state will not be physically explicable. If the difference between our epistemic situation and that of the zombies stems from a difference between the way we are related to our phenomenal states and the way our twins are related to their schmenomenal states, in turn creating a difference in the strength of justification of our corresponding beliefs, then the epistemic relation that holds between a conscious subject and her phenomenal states *cannot be a fully physical relation*, else the zombie would bear the same relation to her schmenomenal states, reuniting our epistemic situations.

If being "acquainted" with our phenomenal states is an essential part of our epistemic situation, and the acquaintance relation is not a fully physical relation, then our epistemic situation cannot be understood in topic neutral terms. In making acquaintance an essential part of our epistemic situation, the dualist will have stipulated that our epistemic situation contains a relation *that can only be understood as (at least partially) non-physical*. If the property dualist is to establish a difference between our epistemic situation and that of the zombie, our epistemic situation cannot be understood in topic-neutral terms.

We may be able to conceive of a scenario in which we do not share our epistemic situation with the zombie, a scenario in which the difference between our epistemic situations is grounded in differing strengths of justification for our corresponding phenomenal and schmenomenal beliefs, the first underwritten by “acquaintance” and the second underwritten only by an ordinary causal connection. But this scenario is conceivable *only if* we assume that it is *conceptually impossible* for both of the relata of the crucial epistemic relation- the one that justifies our beliefs more strongly than the zombies’ can be justified- to be *physical*. If the epistemic inequality of our phenomenal and schmenomenal beliefs stems from a difference in the strength of justification supporting such beliefs, then our epistemic inequality is conceivable *only if we assume that physicalism is false*. It looks as though the physicalist has no reason to accept any of the three preceding property dualist arguments against the epistemic equality of corresponding human and zombie beliefs.

4. The Schmenomenal-Zombie: Friend or Foe? Re-evaluating the “Schmenomenal” Concept Strategy

4.1 False Beliefs about False Consciousness

The property dualist may not be able to argue for an epistemic difference between phenomenal and schmenomenal beliefs by appealing to a form of justification that can only apply to beliefs about *non-physical* phenomenal properties (as the argument from the special epistemic relation of “acquaintance” seems to do), nor can he appeal to functional differences between our phenomenal knowledge and whatever corresponding knowledge our zombie-twins will have. And the physicalist has no reason to accept that the purported distinction between the “richness” of our introspectable phenomenal states and the character of a zombie’s functionally-corresponding, potentially introspectable, states entails a difference in the epistemic status of their beliefs. But the failure of these lines of argument does not guarantee the success of a Schmenomenal Concept Strategy: humans and schmenomenal-zombies may not share their epistemic situation, even if schmenomenal-zombies successfully refer to their own schmenomenal states with their own schmenomenal-concepts. Our zombie-twins’ may have *false* beliefs about a *false* notion of consciousness.

False Beliefs About *False Consciousness*:

Zombie-“phenomenal” concepts differ from human phenomenal concepts in that they neither refer, nor purport to refer, to phenomenal properties (i.e., the referents of our phenomenal concepts).

Zombie-“phenomenal” concepts refer to properties that exist in the zombie world and have the physical, functional, and intentional features of corresponding human phenomenal properties, yet *some of your zombie-twin’s “schmenomenal” assertions may be false when your corresponding phenomenal assertions are true.*

This less-optimistic interpretation of the relationship between corresponding zombie and human beliefs may be counter-intuitive, and likely unpopular among those who would argue that zombies are both conceivable and share our epistemic situation. The zombie-“phenomenal” concept still refers to “schmenomenal” properties, defined as they are above. And- as always- your zombie-twin’s behavior is molecule-for-molecule identical to your own. But this pessimistic interpretation allows for the possibility that humans and their zombie-twins may fail to share their epistemic situations. Attractive or not, this interpretation does adhere to the *letter* of the “schmenomenal-concept” account of zombie-“phenomenal” beliefs, it merely betrays its spirit. As such, it is a potentially coherent interpretation of the zombie’s epistemic situation, and should be investigated.

If the property dualist were to allow the zombie’s “phenomenal” concept to have an alternative- non-phenomenal- referent, he would certainly be more amenable to this interpretation; at the very least, it leaves the epistemic status of schmenomenal beliefs an open question.⁹⁶ As Chalmers writes,

One can develop this intuitive consideration [that zombie-“phenomenal” beliefs are not epistemically identical to their twins’ corresponding phenomenal beliefs] by considering a zombie’s utterances of sentences such as “I am phenomenally conscious.” It is not clear exactly what a zombie asserts in asserting this sentence. But it is plausible that the zombie does not assert a truth. (Chalmers, 2007: 177)

⁹⁶ The official property dualist position is that there is no replacement zombie-state to which the zombie’s corresponding “phenomenal” concepts refer.

Allowing the zombie-phenomenal concept a referent of its own creates logical space for your zombie's "schmenomenal" beliefs to have the same truth value as your phenomenal beliefs, but it does not guarantee this conclusion. There may be cases in which you speak truly while your zombie counterpart- still deploying schmenomenal-concepts and successfully referring to schmenomenal properties- utters identical phonemes, but his statement is false. Given that you and your zombie-twin are functionally identical, it is hard to see how your twin could possess a concept that successfully refers to some non-phenomenal zombie-property (one he actually possesses), a concept that composes zombie-phenomenal beliefs and plays the same role in his cognitive economy as your phenomenal concept and belief play in yours, and yet when he utters all the sounds you utter when you make claims related to phenomenal consciousness, the nature of your twin's concept and its referent may be such that he can speak falsely when you speak truly.⁹⁷ But nothing we have said thus far about phenomenal properties or the content of zombie-"phenomenal" beliefs has ruled out this odd possibility; we need more information about ersatz-zombie-referents and the nature of these corresponding zombie-"phenomenal"-concepts before we can make a legitimate ruling on this matter.

4.1a What the Schmenomenal-Zombie Knows

If my phenomenally-bereft twin is a *schmenomenal*-zombie, then when I reason about scenario S, which is partially composed of phenomenal concepts, my twin is reasoning about scenario Ssz. The only difference between Ssz, as contemplated by my schmenomenal-

⁹⁷ See also Dennett's (1991: 310-311) account of the "the zimboe"- a philosophical-zombie-character who would hold just such false beliefs (that is, by this popular criteria for truth and falsity of zombie-beliefs).

zombie twin, and S, as contemplated by me, is that in place of the phenomenal concepts that appear in S, the corresponding zombie-“schmenomenal”-concepts will appear in Ssz, concepts that refer to states that exist in the zombie-world and which the zombie does possess, states that have the same physical, functional, and intentional properties as phenomenal states but have no *phenomenal feel*. My schmenomenal-zombie-twin’s *expression*⁹⁸ of scenario Ssz will be phonetically identical to my expression of scenario S, and Ssz will be as intelligible to her as S is to me. In the same vein, her expression of her judgment about the conceivability of Ssz will be as intelligible to her as my judgment about the conceivability of S is to me.

A schmenomenal-zombie will judge that she has phenomenally conscious experiences, and if we assign schmenomenal content to her “phenomenal” concept, the truth value of her assertion will correspond to that of her human-twin: both should come out true. Per Carruthers’s and Veillet’s (2007) example, schmenomenal-Zombie Chalmers will judge that he is phenomenally conscious, and he will be “correct when he says that he is conscious,” because his statement is about schmenomenal consciousness. Schmenomenal-Zombie Chalmers,

isn’t saying that he has phenomenal states as we understand them. He is correct because he means that he has schmenomenal states, and he has them. (2007, 231)

Of course, schmenomenal-Zombie Chalmers would vehemently object to the accusation that *his* concept ‘*phenomenal*’ means “schmenomenal”. Schmenomenal-Zombie Chalmers would fight tooth and nail against the claim that he intends (implicitly or explicitly)

⁹⁸ Referring back to the distinction between “intelligibility” and “conceivability”

that his phenomenal concepts refer to schmenomenal states⁹⁹. Zombie Chalmers may be correct when he says he has phenomenal states, but he has no epistemic access to the *reason* his judgment is correct.

The schmenomenal-zombie cannot know that her “phenomenal” concept is a *schmenomenal-concept*; she has no epistemic access to the fact that her “phenomenal concept” has schmenomenal content rather than phenomenal content, that her zombie-“schmenomenal” concepts have only physical, functional, and intentional content. Her inability to know the difference between phenomenality and schmenomenality- her lack of epistemic access to this difference- is vital when she is considering the conceivability of a scenario involving “phenomenal” concepts.

4.1b Inverted-Schmenomenal-Zombie-Twins

If it is conceivable that zombies share their human-twins’ epistemic situation, then epistemically identical “Inverted-Twins” should also be conceivable: Molecule-for-molecule physically-identical subjects, each of whom has a belief or state or concept that is “inverted” relative to the other’s.¹⁰⁰ These “Inverted-Twins” will have appropriately corresponding beliefs about corresponding experiences; the relevant content of both the belief and the

⁹⁹ Chalmers (1996) suggests that we should take the zombie-dualist “at his word”. There is some sense to this claim, but unfortunately there is more than one sense- is Zombie-Chalmers’s “word” to be interpreted as being about phenomenal states, as we know them, or about schmenomenal states, as zombies supposedly know them. “Taking the zombie at his word” does not help clarify how we should interpret the zombie’s “phenomenal” statements.

¹⁰⁰ There are physicalists who hold that inverted *functional*-twins are conceivable, i.e., subjects who share all their functional properties but who differ in their phenomenal experience and subsequent phenomenal beliefs. Oddly, most see no problem with admitting to the conceivability of molecule-for-molecule identical zombie-human twins, one of whom has no phenomenal experience whatsoever. Perhaps these physicalists actually believe zombies are only *prima facie* conceivable, or that zombie-scenarios are intelligible, but would not go so far as to say zombies were ideally conceivable.

experience is “inverted” relative to that of their twin’s. For any phenomenally conscious subject it is conceivable that they have a phenomenally-inverted twin; Brian and Ryan, with their inverted color experiences and phenomenal-color beliefs were our first example. Mary the Color-Scientist has an “inverted” twin as well- call her “Inverted Mary”.

Since we have stipulated the contents of the schmenomenal-zombie’s concepts and beliefs, we can determine whether her beliefs are true or false based on the content of her corresponding human-twin’s phenomenal belief.¹⁰¹ And we know the content of Mary’s and Inverted Mary’s phenomenal beliefs- these contents have also been stipulated. When Mary and Inverted Mary first perceive color, they will acquire epistemically identical beliefs with different phenomenal content, a situation analogous to Oscar’s and Twin-Oscar’s epistemic situation regarding (t)water beliefs.

Schmenomenal-zombies of both the dualist and a posteriori physicalist persuasion will judge that “zombie-scenarios” and “invert-scenarios” are conceptually consistent, that they can conceive of zombies and inverts. The schmenomenal-zombie will judge that the content of her “phenomenal” concepts (schmenomenal-concepts) and the relations between those and the other pertinent concepts allows her to conceive of two phenomenally (schmenomenally) conscious subjects whose physical and functional features are identical to each other but whose “phenomenal” (schmenomenal) experiences are inverted relative to one another.

But we have stipulated the content of the schmenomenal-zombie’s concepts: for every human phenomenal state or concept, a schmenomenal-zombie has corresponding

¹⁰¹ This will certainly work if the content of her human-twin’s belief is also stipulated.

schmenomenal-concepts and schmenomenal states with “the same physical, functional, and intentional properties as [phenomenal] states, but that aren’t phenomenally conscious” (Carruthers and Veillet, 2007: 9-10). When we have beliefs about our occurrent phenomenally conscious states, or general beliefs about phenomenal consciousness, our schmenomenal-zombie-twins will be, “thinking about schmenomenal consciousness using their phenomenal [schmenomenal] concepts” (ibid: 10), and thinking about their occurrent schmenomenal state using their schmenomenal-concepts.

When a schmenomenal-zombie believes she is contemplating “phenomenal inversion”, she is actually thinking about *schmenomenal inversion*. “Phenomenal inversion” is a case in which subject A and subject B are *identical* in all their physical, functional, and intentional properties while their phenomenal properties are *inverted* relative to one another. A corresponding “schmenomenal inversion” would be a case in subject A and subject B are identical in all their physical, functional, and intentional properties while their *schmenomenal properties* are inverted relative to one another.

But of course schmenomenal properties just are the physical, functional, intentional, and otherwise *non-phenomenal* content of (Chalmers’s version of) a phenomenal state. It follows that “schmenomenal inverts” would be two subjects, A and B, whose non-phenomenal properties (physical, functional, and intentional, etc.) are *identical*, while their physical, functional, and intentional properties are *inverted* relative to one another. But a set of properties cannot be both *identical* and *inverted* in the same way at the same time, as the definition of “Schmenomenal-Zombie-Inverted-Twins” would require. Thus the notion of a “Schmenomenal-Zombie-Inverted-Twin” is internally inconsistent, self-contradictory, conceptually incoherent, and inconceivable. The content of the schmenomenal-zombie’s

schmenomenal-concepts, schmenomenal-beliefs, and schmenomenal-states make “Qualia Inverts” and “Qualia Inversion” actually *inconceivable for the schmenomenal-zombie*. The notions of (schmenomenal-zombie)Inverted-Twins and (schmenomenal-zombie)Zombies are conceptually incoherent, despite the dualist and a posteriori physicalist schmenomenal-zombie’s arguments to the contrary.

We know that the content of the schmenomenal-zombie’s schmenomenal-concept would not allow her to actually conceive of “Inverted-Twins”. The schmenomenal-zombie *judges falsely*- she will be wrong about the content of her concepts and wrong about the relation between them. Despite the dualist- and a posteriori physicalist-schmenomenal-zombies’ arguments to the contrary, the content of the schmenomenal-zombie’s schmenomenal-concepts make “Qualia Inverts” and “Qualia Inversion” *inconceivable* for the schmenomenal-zombie. But according to their human-twins, the property dualists and a posteriori physicalists, the statement, “It is conceivable for the functional, physical, and intentional features of a subject’s phenomenal experience to remain unchanged while the phenomenal quality of that experience changes from phenomenal-red to phenomenal-blue,” is made *true* by the meaning of our phenomenal concepts. So on the “schmenomenal” interpretation of zombie-“phenomenal” concepts, our beliefs about the conceivability of zombies and our schmenomenal-zombie-twins’ corresponding schmenomenal beliefs *do not share their truth value*- a problem for the schmenomenal concept strategy.

4.2 Epistemically Equivalent Beliefs, or Merely Epistemically Similar?

Up to this point, our criteria for “sharing an epistemic situation” have been that subjects’ corresponding beliefs must have corresponding truth value and epistemic status, as justified or unjustified, cognitively significant or insignificant¹⁰². Perhaps these criteria are overly demanding. Could we alter these criteria such that our zombie-twins could share our epistemic situation without being our epistemic-*twins* as well?

Generally speaking, “sharing an epistemic situation” shouldn’t require having epistemically *identical* beliefs. In chapter 3, I gave the example of learning by textbook and by lecture; though their newly acquired beliefs would not be epistemically *identical* (among other discrepancies, their beliefs would have different sources of justification), I suggested that both textbook-reading and lecture-listening could give the reader and listener epistemically *equivalent* knowledge of the relevant subject matter.

Could there be some case for an epistemically appropriate similarity between our phenomenal beliefs and our zombie-twins’ corresponding beliefs that does not require that they have corresponding truth value? The phenomenal concept strategy might be satisfied by a less-than-perfect epistemic symmetry between the zombies’ beliefs and those of their human twins.

For either the Schmenomenal Concept Strategy or general phenomenal concept strategy to succeed, the relation between our phenomenal beliefs and our zombie-twins’ schmenomenal beliefs must be *epistemically “similar enough”* for our zombie-twins to share

¹⁰² -here we were only looking at their potential cognitive significance and conceptual distinctness

our epistemic situation- the distinction between our epistemic situations must be significant; it must be a distinction that makes a difference. If the schmenomenal-zombie may be wrong in some of her beliefs while her human-twin's corresponding belief is correct, we must remove "corresponding truth value" from the criteria for sharing an epistemic situation. Perhaps we can still hold on to the "cognitive significance" requirement. It seems that something in this area needs to remain fixed; otherwise it is quite difficult to see how to defend the claim that zombies share our epistemic situation, a claim that neither the schmenomenal nor general phenomenal concept strategist can let go.

4.2b Investigating the Cognitive Significance of Schmenomenal beliefs

When the notion of "schmenomenal belief" was originally introduced, I said the zombies' schmenomenal beliefs were as significant to the schmenomenal-zombies as our phenomenal beliefs are to us. The zombie's schmenomenal beliefs play as significant a role in the schmenomenal-zombies' cognitive economy as our phenomenal beliefs play in our own. And I believe this is "one good sense" in which a belief can be *cognitively significant*. In this chapter, we will interpret "cognitive significance" in different way. Here we will treat it as a more formal notion, one with explicit criteria for satisfaction that a belief might satisfy or fail to satisfy. This more formal notion of "cognitive significance" can be used in conjunction with other criteria to test the tenability of the Schmenomenal Concept Strategy.

Though the criteria for cognitive significance will vary between theories and contexts, one seemingly necessary criterion for cognitive significance of any kind would that a belief B cannot be cognitively significant if both B and its negation, $\sim B$ (or any belief that is

incompatible with/entails the denial of B) constrain *the same the class* of a priori epistemic possibilities, or eliminate the same class of epistemically possible worlds. That is to say, if there is no *conceptual difference* between B and its negation, then B is not cognitively significant.

This way of determining the cognitive significance of a particular belief might sound odd at first, but it has benefits. The first is that in using these minimalist criteria for judging cognitive significance, we do not need to address the larger questions surrounding cognitive significance; questions about the *real content* of these beliefs, their *actual* meaning, or whether the concepts composing them are empty or actually refer. The second (related) benefit is that they are entirely topic neutral: these criteria judge the cognitive significance of a belief only relative to another belief that is *stipulated* as having the same kind of content, *whatever* that content might (or might not) be.

4.3 Zombies, Inverts, and Schmenomenality- Peacefully Coexistent or Secretly Antagonistic?

Our friends, Mary and Inverted Mary, are stipulated as having *inverted beliefs* about “*what it's like*” to be in particular *experiential*-states, experiential-states that are inverted relative to one another while their functional states are identical to one another (in the Mary story, these experiential and functional states are brought about by seeing a tomato with certain physical properties under certain physical conditions). Their beliefs are stipulated as *independently* cognitively significant (they must be, if their new beliefs constitute “new propositional knowledge” in any way) and cognitively significant *relative to one another's*. If the Schmenomenal Concept Strategy holds, then (given our reduced criteria for “sharing an

epistemic situation”) the corresponding beliefs of Mary’s and Inverted Mary’s schmenomenal-zombie-twins, schmenomenal-Zombie Mary and schmenomenal-Zombie Inverted Mary, should be both independently cognitively significant and cognitively significant relative to one another’s.

If the schmenomenal-zombie- and human-Marys’ corresponding phenomenal and schmenomenal beliefs do not share their epistemic status as cognitively significant or insignificant, then, even by the current, lowered, standards for “sharing an epistemic situation”, the Human-Marys’ epistemic situation will differ from that of the Schmenomenal-Zombie-Marys’. At that point, the “Schmenomenal Concept Strategists” will have several options, one of which is to again reformulate the criteria for sharing an epistemic situation (rejecting the corresponding cognitive significance requirement *as well as* the corresponding truth value requirement). Another option- possibly as a last resort- is to embrace the view that zombie-“phenomenal” talk is meaningless, and disregard the idea of zombies having an “epistemic situation” altogether. Whichever option they choose, none will be as straightforward as giving the zombie-“phenomenal” concept a referent of its own, one that exists in the zombie world, one to which the zombie actually *has* epistemic access (perfect or not), a referent that can make the zombies’ “phenomenal” statements true in virtue of simply “meaning something other than” what our phenomenal statements mean.

4.3a The Four Faces of Mary: Our Cast of Characters

We are already familiar with the notion of an “inverted-twin”, and with the notion of a “zombie-twin”. The first, an inverted-twin, is customarily a human whose phenomenal

belief is “inverted” relative to another human’s phenomenal belief. The second, a “zombie-twin”, is a zombie who is physically and/or functionally identical to a human being. The “Inverted-Twin”-relation typically applies to set consisting of two humans, and the “Zombie-Twin”-relation typically applies to a set consisting of one zombie and one human. These are not the only possible sets to which these relations may apply, and it should be no surprise that these relations, in different combinations, can give rise to problems for the conceivability of some phenomenally-odd but apparently conceptually coherent creatures.

In the case of Mary the Super Scientist, these relations seem to create four characters whose features we will explore below. There is the original human Super-Scientist- here we will call “Ordinary Mary”. Ordinary Mary has an inverted-twin, Inverted Ordinary Mary, herself a human being. Ordinary Mary also has a zombie-twin, Zombie Ordinary Mary, a zombie-version of the human Ordinary Mary. And the human Inverted Ordinary Mary has a zombie-twin as well, Zombie Inverted Ordinary Mary.

So our cast has four Mary members: Ordinary Mary, Inverted Ordinary Mary, Zombie Ordinary Mary, and Zombie Inverted Ordinary Mary. That is to say, our cast begins with four Marys. These zombies are schmenomenal-zombies, and schmenomenal-zombies have some peculiar features; features which may make one of these Mary-notions incoherent. If so, though all four seem conceivable, in the end we will have but three.

4.3a1 Ordinary Mary

Mary, as we know, steps out of her black and white room, sees her first red-colored tomato, acquires a new phenomenal concept and a phenomenal belief which she expresses with the phrase, “This is what it's like to see red.” On Chalmers’s account of phenomenal

epistemology (and in his parlance), Mary's new phenomenal belief is " $E = R$ ", where E is a demonstrative concept, picking out *whatever phenomenal state Mary ended up in when she experienced her perception of red*, and R is a substantive concept, picking out her phenomenal experience itself *as the experience that it is* (or "by its intrinsic nature"), rather than as *whatever experience Mary happened to have* when she perceived the red-colored tomato.¹⁰³ Ordinary Mary's R concept was conceived in the event 'Ordinary Mary perceives her first colored-tomato' and its content is determined independently of any previously acquired concept or belief.

For the purposes of this thought-experiment, we will call Mary's new phenomenal belief " $(TF)E=R$ ", where (TF) represents the functional state that Mary and her inverted twin share in common, E is the phenomenal experience or *what it's like* to have *that experience* (the one that is accompanying functional state (TF)), and R is as defined above. The belief $(TF)E=R$ constrains the class of a priori epistemic possibilities to those compatible with her phenomenal experience and belief; the truth of $(TF)E=R$ eliminates any epistemically possible worlds in which $(TF)E$ is not an R experience.

If $(TF)E=R$ is cognitively significant, then (among other things) the truth of $(TF)E=R$ should eliminate a *different* set of a priori epistemic possibilities than would the denial of $(TF)E=R$ (or a belief incompatible with $(TF)E=R$). As Chalmers writes, "the only epistemic possibilities compatible with her (Mary's) belief are those in which tomatoes cause R experiences" (2003: 232).

¹⁰³ It is an open question whether there is a *real* difference between these two concepts (E and R), and this is a subject worth addressing in itself (a subject I address in other work). For the purpose of this thought-experiment we can leave the question aside.

4.3a2 Inverted Ordinary Mary

When Inverted Ordinary Mary leaves her monochromatic home, she enters into a perceptual state that is molecule-for-molecule identical to that of her inverted-twin, state “TF”. In this perceptual even she acquires a new phenomenal concept and phenomenal belief that share the form Ordinary Mary’s concept and belief, but with different phenomenal content. Inverted Mary’s belief is $E=G$ (where the concepts “ E ” and “ G ” functionally correspond to Mary’s E and R). Let’s call Inverted Mary’s new belief “(TF) $E=G$ ”. Inverted Mary expresses her new phenomenal belief (TF) $E=G$ with the same phrase as Mary expresses (TF) $E=R$: “This is what it’s like to see red.”

(TF) $E=G$ constrains the class of a priori epistemic possibilities to only those compatible with Inverted Mary’s phenomenal experience, i.e., worlds in which tomatoes cause G experiences (in someone molecule-for-molecule identical to Inverted Mary). If (TF) $E=G$ is cognitive significant, the truth of (TF) $E=G$ will constrain a different class of a priori epistemic possibilities than those compatible its denial.

According to these criteria, are Ordinary Mary’s and Inverted Mary’s new phenomenal beliefs cognitively significant? Yes. The set of worlds in which tomatoes cause red experiences in creatures molecule for molecule identical to Mary and the set of worlds in which tomatoes cause green experiences in such creatures are mutually exclusive sets. Mary and Inverted Mary have cognitively significant phenomenal beliefs at least insofar as their beliefs are stipulated as having incommensurable phenomenal content, and insofar as this content is conceptually independent from functional, intentional, and other non-phenomenal content.

Inverted Mary's new phenomenal belief, (TF)E=G, resulting from her 'what it's like to perceive-tomato-M' experience, imposes a different constraint on the class of epistemic possibilities than the constraint imposed by the phenomenal belief (TF)E=R, which resulted from the experience of her inverted-twin, Original Mary. Can the same be said for Original Mary's and Inverted Mary's schmenomenal-zombie-twins, schmenomenal-Zombie Mary and schmenomenal-Zombie Inverted Mary?

4.3a3 Schmenomenal-Zombie Mary

When schmenomenal-Zombie Mary steps out of her black and white room and sees her first tomato (tomato-M), she has thoughts or beliefs about her resulting schmenomenal state that correspond to the thoughts and beliefs that her human-twin, Ordinary Mary, has about her phenomenal state. But schmenomenal-Zombie Mary is thinking her thoughts with schmenomenal-concepts, and forms the schmenomenal belief, '*S-this=S-**', which corresponds to Mary's phenomenal belief, '*E=R*'. Translated, '*S-this=S-**', means something like, '*this schmenomenal state I am currently in is the schmenomenal state I am identifying with the schmenomenal-concept '-**'.¹⁰⁴ Schmenomenal-Zombie Mary expresses her new schmenomenal belief with a phrase phonetically identical to the phrase with which her human-twin expresses her phenomenal belief: "This is what it's like to see red".

¹⁰⁴ The phrase, "the schmenomenal-concept '-*'" is awkward, but it will avoid any unnecessary confusion that would come with using a phrase such as "the schmenomenal-concept 'redness'" to represent schmenomenal-Zombie Mary's new schmenomenal-concept, since a public-language term such as "redness"- or any pronounceable expression, even one as strange as 'plusghr'- can inadvertently lull us into believing we have a prima facie grasp of the content of schmenomenal-concepts. What sort of content a schmenomenal-zombie's schmenomenal-concept would have is actually quite a tricky matter that-for our purposes here- we do not need to address, and can avoid by representing them with the minimalistic '-*' and 'Inverted-*'.

In schmenomenal-Zombie Mary's new schmenomenal belief, the schmenomenal-concept "*S-this*" is a demonstrative concept, functioning just as Ordinary Mary's '*E*', picking out *whatever schmenomenal state is caused in schmenomenal-Zombie Mary in virtue of perceiving tomato-M*.

"S-*" in schmenomenal-Zombie Mary's belief represents the schmenomenal-concept '-*'. 'S-*' functionally corresponds to Ordinary Mary's concept, *R*; it is the concept that schmenomenal-Zombie Mary acquires when she sees tomato-M and goes into the corresponding schmenomenal state. Like Ordinary Mary's phenomenal concept, *R*, schmenomenal-Zombie Mary's concept, 'S-*', picks out her occurrent schmenomenal state *itself, as the schmenomenal state that it is* (unlike her demonstrative concept '*S-this*', which denotes *whatever schmenomenal state schmenomenal-Zombie Mary happens to be in at that moment*). And like Mary's *R* concept it was conceived in the event 'Schmenomenal-Zombie Mary perceives tomato-M', and its content is determined independently of any concept or belief schmenomenal-Zombie Mary has previously acquired.

The features that individuate schmenomenal-Zombie Mary's new schmenomenal state and corresponding schmenomenal belief, *S-this=S-**, or informally, '(TF)=S-*', from *other* schmenomenal states and schmenomenal beliefs is that '(TF)=S-*' has all the physical, functional, and intentional properties as Mary's phenomenal state ('*what it's like* to perceive tomato-M') and phenomenal belief ('(TF)E=R'), but as a schmenomenal state, schmenomenal-Zombie Mary's 'perceiving-tomato-M' and schmenomenal belief, '(TF)=S-*', have no phenomenal content.

4.3a4 Schmenomenal-Zombie Inverted Mary

Schmenomenal-Zombie Inverted Mary is defined by her multiple twin relations¹⁰⁵: she is both the *schmenomenal-zombie-twin* of Inverted Mary and the *inverted-twin* of schmenomenal-Zombie Mary. When Schmenomenal-Zombie Inverted Mary first sees tomato-M, she thinks about her occurrent schmenomenal state, the state that 1) *corresponds with* Inverted Mary's phenomenal state and 2) is *inverted relative to* schmenomenal-Zombie Mary's schmenomenal state. Like schmenomenal-Zombie Mary, Schmenomenal-Zombie Inverted Mary thinks her thoughts with her schmenomenal-concepts, and uses the phrase "This is what it's like to see red," to express her newly acquired schmenomenal belief, '*S-this=S-Inverse**' (translated: '*this schmenomenal state I am currently in is the schmenomenal state I am identifying with the schmenomenal-concept 'Inverse**').¹⁰⁶

Schmenomenal-Zombie Inverted Mary's schmenomenal concept, the "*S-this*" is a demonstrative concept that behaves just as schmenomenal-Zombie Mary's "*S-this*" behaves, picking out *whatever* schmenomenal state Schmenomenal-Zombie Inverted Mary is in at that moment. The "*S-Inverse**" in Schmenomenal-Zombie Inverted Mary's belief is the schmenomenal-concept, *Inverse**; which corresponds to Inverted Mary's phenomenal concept *G*. Schmenomenal-Zombie Inverted Mary's belief, '*(TF)=S-Inverse**', corresponds to Inverted Mary's phenomenal belief, '*(TF)E=G*'; it has all the physical, functional, and intentional properties as '*(TF)E=G*', but as a schmenomenal belief, '*(TF)=S-Inverse**', has none of the phenomenal content.

¹⁰⁵ The other three characters in this thought-experiment are as well, but the point is worth stressing here.

¹⁰⁶ It is better to use "schmenomenal-concept*" than "schmenomenal-redness" to denote the schmenomenal-concept schmenomenal-Zombie Mary is deploying to refer to her schmenomenal state, since the content of schmenomenal-concepts is a tricky matter that- for our purposes here- we do not need to address.

In virtue of being the schmenomenal-zombie-twin of Inverted Ordinary Mary, Schmenomenal-Zombie Inverted Mary's belief 'S-Inverse*' should bear the same "inversion" relation to schmenomenal-Zombie Mary's '-*' as Inverted Ordinary Mary's concept G bears to Ordinary Mary's concept R, i.e., it should be *inverted relative to* the content of schmenomenal-Zombie Mary's schmenomenal-concept '-*'.

4.4 "Mary-Relations": The Deceptive Intelligibility of "Correspondence" and "Inversion"

Now that we have laid out the contents of the four "Marys" phenomenal and schmenomenal beliefs in explicit detail, the peculiarities of the relations between these characters and the interrelated features of their beliefs are starting to become clear. One concerns the cognitive significance of the two schmenomenal-zombie Marys' schmenomenal beliefs: are the schmenomenal-zombie's newly acquired schmenomenal beliefs cognitively significant relative to one another's?

We can address the question of whether schmenomenal-Zombie Mary's and Schmenomenal-Zombie Inverted Mary's new schmenomenal beliefs are cognitively significant by applying the same "comparative" criteria that affirmed the cognitive significance of Ordinary Mary's and Inverted Ordinary Mary's new phenomenal beliefs;¹⁰⁷ i.e., comparing the content of schmenomenal-Zombie Mary's newly-acquired schmenomenal belief to that of her inverted-twin, and seeing whether they eliminate different sets of epistemically possible worlds.

¹⁰⁷ To be more precise, their phenomenal beliefs are cognitively significant insofar as we assume that these Marys form new phenomenal beliefs when they see tomato-M, and insofar as their beliefs are stipulated to be conceived of as having distinct phenomenal content.

The stipulated definition of the Schmenomenal-Zombie, along with the stipulated definition of Mary and Inverted Mary, should give us all the information we need to determine the relation between schmenomenal-Zombie Mary's and Schmenomenal-Zombie Inverted Mary's schmenomenal beliefs. Each zombie's schmenomenal belief is stipulated as sharing its epistemic status with her corresponding human-twin's phenomenal belief, and these zombies are stipulated as the zombie-twins of "inverted" human-twins. Since we know the relation between Ordinary Mary's and Inverted Ordinary Mary's phenomenal beliefs, '(TF)E=R' and '(TF)E=G'- they are conceptually distinct and cognitively significant relative to each other- their schmenomenal-zombies' corresponding schmenomenal beliefs ought to be conceptually distinct and cognitive significant relative to one another's.

"Zombies", "inverted-twins", and "Schmenomenality" are all intelligible notions and seem to be conceptually compatible. Prima facie, it seems conceivable the four Marys could be epistemically related in this way. But on reflection the postulated epistemic symmetry of the schmenomenal-zombie beliefs and the phenomenal beliefs of their "inverted" human twins will not hold; Schmenomenality collapses under the weight of scrutiny, and takes the Schmenomenal Concept Strategy with it. Applying the "Schmenomenality"-interpretation of zombie-beliefs to these four Marys creates chimerical "zombie" creatures with competing, internally inconsistent features. To see why, we need to look more closely at the relationship between each of these four "Marys".

4.4a Inverted-Twinship

To piece together the case for this collective incompatibility, we can begin by looking at the essential features of the multiple "Mary-Relations", starting with the "inverted-

twins” of Ordinary Mary and Inverted Ordinary Mary.¹⁰⁸ A set of “Inverted-twins” are a pair of subjects, Twin A and Twin B, who are identical in some specified respects,¹⁰⁹ while Twin A has some specified state or belief that is “inverted” relative to (or in some appropriate way, incompatible with) a corresponding state or belief belonging to Twin B. By definition, inverted-twins must have at least one pair of conceptually distinct corresponding beliefs (the content of which will put different constraints on the set of a priori epistemic possibilities relative to the content of the other);¹¹⁰ these beliefs will be cognitively significant relative to one another.

4.4b Human / Schmenomenal-Zombie-Twinship

The essence of the “Human-Schmenomenal-Zombie-twin” relation is in the epistemic equivalence between corresponding Schmenomenal-Zombie schmenomenal properties and human phenomenal properties. Your Schmenomenal-Zombie-twin’s schmenomenal properties have “the same physical, functional, and intentional properties as [phenomenal] states”, but they lack *phenomenality*. In its place, these states have *schmenomenality*, which is “epistemically just as good as” phenomenality, but has no *phenomenal feel*. You’re Schmenomenal-Zombie-twin’s schmenomenal-concepts successfully refer whenever your corresponding phenomenal concepts successfully refer, and his schmenomenal beliefs are true whenever your corresponding phenomenal beliefs are true (false when they are false).

¹⁰⁸ This seems obvious: what makes Inverted Mary *Inverted Mary* is that her phenomenal belief (and phenomenal experience) is inverted relative to Mary’s, and a minimal criterion for a belief T-R being “inverted” compared to belief T-G is that *T-R is not the same as T-G*.

¹⁰⁹ They are usually “phenomenal-inverts”, sharing all their non-phenomenal properties in common.

¹¹⁰ Or at least they are stipulated *to be conceived of as* distinct; conceiving of them in any other way would violate this stipulated definition of “inverted-twin”.

But his concepts refer to *different* properties and states than those to which your phenomenal concepts refer. Your schmenomenal-zombie-twin's statement, "I have phenomenally conscious states," is true, because he is using the term "phenomenally" to mean "schmenomenally"; his statement is true because it is about your schmenomenal-zombie-twin's schmenomenal states, states he does possess.

4.5 The Results are In, and Odd:

What is the essential feature of the relation between the zombie-zombie-twins, Schmenomenal-Zombie Mary and Schmenomenal-Zombie Inverted Mary? One uncontroversial aspect of the relation between the two "Zombie-Marys" is that each is the zombie-twin of one member of a pair of human "inverted-twins". Beyond this, the waters get murky. Zombie-Twins are easy to conceive. We all have our zombie-twins: Ordinary Mary has a zombie-twin, and Inverted Ordinary Mary has a zombie-twin as well. "Inverted-twins" are as easily conceivable as Ordinary Mary and Inverted Ordinary Mary. And "Schmenomenality" is simply one among many apparently coherent interpretation of the epistemic relation between the zombie's epistemic situation and our own.¹¹¹

Given the apparent conceptual coherence of these three notions- Inverted-twins, Zombies, and Schmenomenality- the notion of "Inverted Schmenomenal-Zombie-Twins" should not be difficult to conceive. Yet the relation between these twins, as both Schmenomenal-Zombies and as the zombie-twins of a pair of human-inverted-twins, seem to

¹¹¹ "Schmenomenality", again, is the interpretation of zombie "phenomenal" concepts and beliefs according to which zombie "phenomenal" (schmenomenal) concepts refer to particular zombie-properties and states that are epistemically "just as good" as their human-twins', and that the truth value and cognitive significance of a zombie's "phenomenal" concepts and beliefs correspond with their human twins' counterpart phenomenal concepts and beliefs, and that we share our epistemic situation with the zombies.

impose conflicting demands on the properties of these Schmenomenal-Zombies' and the content of their schmenomenal beliefs.

According to the Schmenomenal-concept strategy, the content of both of the Schmenomenal-Zombie Marys' schmenomenal beliefs share the epistemic features of their respective Human-Mary-Twins' *inverted* phenomenal beliefs. Ordinary Mary's phenomenal belief and Inverted Ordinary Mary's phenomenal belief are cognitively significant relative to one another's. This "relative cognitive significance" is an epistemic feature of the Human-Mary-Twins' inverted phenomenal beliefs, so the Human-Marys' zombie-twins' beliefs ought to have this epistemic feature as well. The content of Zombie Ordinary Mary's schmenomenal belief and Zombie Inverted Mary's schmenomenal belief ought to be cognitively significant relative to one another. And each of the Zombie-Marys' relations to her corresponding member of the "Human-Mary-Inverted-Twin" pair seem to demand that the content of the zombie/zombie-twins' corresponding schmenomenal beliefs be "inverted" relative to one another.

But these "inverted" zombie-twin's corresponding schmenomenal beliefs cannot be cognitively significant relative to the other, nor can these schmenomenal beliefs be conceptually distinct from each other. Neither can the contents of their schmenomenal beliefs be "inverted" relative to one another's. The demands that the Schmenomenal-concept strategy put on the Zombie Marys' beliefs cannot be met; for the Schmenomenal concept strategy to succeed, the Zombie-Mary-Twins' schmenomenal beliefs must have particular epistemic features that it would be logically impossible for such beliefs to have.

4.5a The Relative Cognitive *Insignificance* of Inverted Schmenomenal-Zombie Beliefs

What constraints will the truth of Schmenomenal-Zombie Mary's schmenomenal belief, '(TF)=S-*', put on the class of a priori epistemic possibilities *relative to* the class constrained by Schmenomenal-Zombie Inverted Mary's schmenomenal belief, '(TF)=S-Inverse*'? We do not need to know the precise content of Schmenomenal-Zombie Mary's belief to determine the set epistemically possible worlds eliminated by '(TF)=S-*'. Since this schmenomenal belief corresponds to Ordinary Mary's phenomenal belief, '(TF)E=R', we know that '(TF)=S-*' has all the physical, functional, and intentional properties of '(TF)E=R' and is true whenever (TF)E=R is true, false when (TF)E=R is false (and as a schmenomenal belief '(TF)=S-*' has no phenomenal content, no associated *phenomenal feel*). In virtue of Schmenomenal-Zombie Mary's Schmenomenal-Zombie-twinship with Ordinary Mary, we know that the truth of '(TF)=S-*' will eliminate all epistemically possible worlds except those consistent with the physical, functional, and intentional properties of Mary's phenomenal belief, '(TF)E=R'. The features that individuate '(TF)=S-*' from *other* schmenomenal beliefs is that it has *the same physical, functional, and intentional properties as Ordinary Mary's phenomenal belief 'what it's like to experience seeing tomato-M'*.

In the same way, we can determine content and the class of epistemically possible worlds Schmenomenal-Zombie Inverted Mary's '(TF)=S-Inverse*' would eliminate by its relation to Schmenomenal-Zombie Inverted Mary's human-twin, Inverted-Ordinary-Mary. '(TF)=S-Inverse*' will eliminate as epistemic possibilities all worlds except those consistent with the physical, functional, and intentional properties of Inverted-Ordinary-Mary's phenomenal belief, '(TF)E=G', which is by definition conceptually distinct from and

cognitively significant relative to Ordinary Mary's belief, '(TF)E=R'. The features that individuate '(TF)=S-Inverse*' from *other* schmenomenal beliefs is that it has *the same physical, functional, and intentional properties as Inverted-Ordinary-Mary's phenomenal belief 'what it's like to experience seeing tomato-M', i.e., '(TF)E=G'.*

The exegesis of the Multi-Mary situation has been tedious; the written-representations of the four Marys' beliefs have become overly familiar, appearing far too many times in these past several pages. But the devil of inconsistency is in a scenario's details, and no argument for the conceptual *incoherence* of an *apparently intelligible* notion will be compelling unless it can show that the devil is undeniably present. Irritating repetition aside, I would guess (or hope) that when the you saw these written representations of the four Marys' beliefs (over, and over) you found them intelligible; perhaps not at first but more so on each encounter. And I would guess (hope) that you found the stipulated relations between the Marys' beliefs intelligible as well. So I hope the problem facing the schmenomenal concept strategy is as painfully obvious as my description of the Marys' beliefs was tedious: is Schmenomenal-Zombie Mary's belief '(TF)=S-*' conceptually distinct and cognitively significant relative to Schmenomenal-Zombie Inverted Mary's belief '(TF)=S-Inverse*'? No. The individuation conditions for '(TF)=S-*' and '(TF)=S-Inverse*' are identical: The features that individuate the schmenomenal belief '(TF)=S-*', are that it has all (and only) *physical, functional, intentional, and otherwise non-phenomenal properties* as Ordinary Mary's belief '(TF)E=R'. And the physical, functional, intentional, and otherwise non-phenomenal properties of '(TF)E=R' are by hypothesis identical to the physical, functional, intentional, and otherwise non-phenomenal properties of Inverted Ordinary Mary's belief '(TF)E=G', the features of '(TF)=S-Inverse*' that individuate it from other schmenomenal beliefs. '(TF)=S-*

and '(TF)=S-Inverse*' have the same individuation conditions- as Ordinary Mary's and Inverted Ordinary Mary's zombie-twins, Schmenomenal-Zombie Mary and Schmenomenal-Zombie Inverted Mary share all their non-phenomenal properties in common, which constitute the entirety of a Schmenomenal-Zombie's properties.

In virtue of sharing their epistemic situation with their human-(inverted)-twins, the contents of Schmenomenal-Zombie Mary's and Schmenomenal-Zombie Inverted Mary's schmenomenal beliefs should be "*inverted*" relative to one another. As Schmenomenal-Zombies, their schmenomenal beliefs are stipulated as sharing their truth value and their epistemic status (as cognitive significant and conceptually distinct) with the phenomenal beliefs of their respective human-twins.

At the same time, in virtue of being *zombies*, the content of Schmenomenal-Zombie Mary's schmenomenal belief and Schmenomenal-Zombie Inverted Mary's schmenomenal belief must be *identical*; their states, concepts, and beliefs have *only* the physical, functional, intentional, and otherwise non-phenomenal content of their human-twins' phenomenal states, concepts, and beliefs.

The features of inverted human-twin beliefs that we are to conceive of as *inverted* is precisely the content that their zombie-twins' schmenomenal beliefs *lack*- phenomenal content. The limitations on the content of Schmenomenal-Zombies' concepts and beliefs (and other properties)- stipulated as physical, functional, intentional, and *all but phenomenal properties*- would make it conceptually incoherent (logically impossible) for any of Schmenomenal-Zombie Mary's and Schmenomenal-Zombie Inverted Mary's properties to differ in any way; not in the content of their schmenomenal-concepts, their schmenomenal beliefs, or their schmenomenal states. There is no logically possible change in

Schmenomenal-Zombie Mary that is not a change in her physical, function, intentional, non-phenomenal properties. So it is conceptually incoherent to “change (or invert) the content of Schmenomenal-Zombie Mary’s schmenomenal state or belief while leaving all the functional etc. facts about Schmenomenal-Zombie Mary the same.”

Given that Inverts, Zombies, and Schmenomenal states are all supposedly coherent notions, it is surprising that combining the three leaves Inverted-Schmenomenal-Zombie-twins with contradictory properties (i.e., leaves the notion of Inverted-Schmenomenal-Zombie-twins internally inconsistent). There is *no conceivable difference- no logically possible difference-* between “Schmenomenal-Zombie Mary” and “Schmenomenal-Zombie Inverted Mary”. The only properties that Schmenomenal-Zombie Mary and Schmenomenal-Zombie Inverted Mary do have, including the content of their schmenomenal beliefs, will be identical.

4.6 The Inverted-Schmenomenal-Zombie-twin: a Genuinely Inconceivable Mind?

I think it is clear that “Inverted-Schmenomenal-Zombie-Twins” are conceptually incoherent, that there is no logically possible difference between “Schmenomenal-Zombie Mary” and “Schmenomenal-Zombie Inverted Mary”. But there is a conceivable difference between Mary and her Schmenomenal-Zombie-twin, Zombie Mary. And there is a conceivable difference between the “Inverted-twins”, Ordinary Mary and Inverted Ordinary Mary. And it is conceivable for each of these Human-Marys to have schmenomenal-zombie-twins: schmenomenal-Zombie Mary and Schmenomenal-Zombie Inverted Mary. But there is

no conceivable difference between Schmenomenal-Zombie Mary and Schmenomenal-Zombie Inverted Mary.

1. Ordinary Mary is conceivable, so Schmenomenal-Zombie Ordinary Mary is conceivable.
2. Inverted Ordinary Mary is conceivable, so Schmenomenal-Zombie-Inverted Ordinary Mary is conceivable.
3. By hypothesis, Ordinary Mary and Inverted Ordinary Mary do not hold identical beliefs, and as such are not qualitatively identical.
4. According to the definition of a zombie, in conjunction with the Optimistic Schmenomenal Concept Strategy, Schmenomenal-Zombie Ordinary Mary and Schmenomenal-Zombie Inverted Ordinary Mary should not be identical, because “Optimistic Schmenomenality” would require that the corresponding schmenomenal-beliefs of these two schmenomenal-zombies differ in their content.
5. But there is no possible difference between Schmenomenal-Zombie Ordinary Mary and Schmenomenal-Zombie Inverted Ordinary Mary.
6. So the Ordinary Marys do not have *distinct* schmenomenal-zombie-twins, contra to what is required by the Optimistic Schmenomenal Concept Strategy

We started with four *conceivably distinct* Mary characters, but we are left with only three. However we want to describe the relationship between our phenomenal beliefs and our zombie-twins’ counterpart-beliefs, “Optimistic Schmenomenality” is an inadequate interpretation of the relation between our phenomenal beliefs and our zombie-twins’ “phenomenal” beliefs. This bears out in the non-distinctness of the Schmenomenal-Zombie-Twins of the human *inverted*-twins, Ordinary Mary and Inverted Ordinary Mary. Positing “schmenomenal properties” as the referent of zombie-“phenomenal” concepts seemed to open a route to epistemic equality, but it led the way to a genuinely inconceivable mind: the Inverted-Schmenomenal-Zombie-Twin.

4.6a Epistemically-Quite-Similar Beliefs?

The relation between phenomenal beliefs and our zombie-twins' corresponding beliefs is much more complicated than it first appears, and nothing we have said so far has resolved the issue of the zombies' epistemic situation. Zombie-“phenomenal” beliefs *correspond* to phenomenal beliefs (in whatever way we want to interpret “correspondence”), but it is still not apparent how the epistemic status of the former compares to the epistemic status of the latter.

We could try for “epistemic equality” rather than “identity”; zombie-“phenomenal” concepts or beliefs may have appropriateness conditions though their assertions are not truth evaluable. Your belief “I am having an experience of phenomenal-redness” may be true, and while your zombie-twin's corresponding expression may not be truth apt, we might be able to argue that it is *appropriate* for her to utter that sentence in a context physically identical to mine.

But it's harder than it seems to determine appropriateness conditions for deploying one zombie “phenomenal” concept rather than another. Even if we allowed appropriateness conditions for deploying zombie-phenomenal concepts, rather than requiring that they have truth conditions, how would these appropriateness conditions be determined? The correctness conditions for deploying a human phenomenal concept seem to have no bearing on the appropriateness conditions for deploying a zombie-phenomenal concept. By hypothesis, there is nothing about the *zombie's state* that makes deploying the concept “*phenomenal-red*” more appropriate in situation R than “*phenomenal-green*”, even though- so long as we make a conceptual distinction between the content of zombie and human mental states- it would be incorrect for me to deploy my phenomenal concept, “*phenomenal-green*”, in an R situation.

And we would still need to answer the question of whether we can “share our epistemic situation with our zombie-twins” in any significant sense if our phenomenal and schmenomenal beliefs share many- but not all- of their epistemic features.

The phenomenal and “phenomenal” beliefs of human and their zombie-twins may be *epistemically-quite-similar*, but can similarity without epistemic identity be *good enough* to show that the a posteriori physicalist can give a fully physicalist account of our epistemic situation regarding the explanatory gap? And when it comes to the content of phenomenal belief, is there really any such thing as being in *epistemically similar but not epistemically identical* situations? These questions highlight the difficulty of arguing that our zombie-twins share our epistemic situation, but are intentionally left open for the reader’s consideration. Perhaps the phenomenal concept strategist can develop plausible criteria for “sharing an epistemic situation” that accommodates these apparent difficulties, or can find a way to dismiss the relevance of the relation between our epistemic situation and that of our zombie-twins’. In either case, the challenge facing the phenomenal concept strategist will be significant.

5. The Conceiver Stays in the Picture: Why “Ideal Conceivability” is a Less-Than-Perfectly-Rational Notion

5.1 Carving up “Conceivabilities” to Explain Away the Appearance of Subjectivity

“Ideal conceivability” is a fundamental assumption of CP; without a rational notion of conceivability there is no Conceivability/Possibility principle, no hope of a link between the epistemic and the metaphysical. As Chalmers writes,

...if we are looking for a notion of conceivability such that conceivability tracks possibility perfectly, we must focus on ideal conceivability. In this sense conceivability is not a merely psychological notion; it is a *rational* notion, in much the same way that a priority and rational entailment are rational notions. If there is to be a plausible epistemic/modal bridge, it will be a bridge between the rational and modal domains. (2002: 160 emphasis original)

Facts about epistemic agents are irrelevant to a priority and entailment: the a priori truth of a proposition is not relative to the subject evaluating the proposition; neither is entailment relative to a subject’s beliefs about the relation between propositions. Though conceivability is in some cases “relative to a speaker or thinker” (2002: para 7), for CP to hold- for conceivability to entail possibility- there must be some variety of conceivability according to which the conceivability of S is not relative to an epistemic agent’s beliefs, a variety of conceivability that is a rational notion, rather than a psychological or subjective notion.

Carving conceivability into several distinct notions gives the property dualist the opportunity explain the appearance of subjectivity in conceivability-judgments; for example, if there are several types of conceivability, it might be plausible that only certain varieties are subject-dependent. The more finely we discriminate between “conceivabilities”, the more plausible it will seem that, while the truth of some types of conceivability-judgments are clearly relative to a conceiver (prima and secunda facie conceivability-judgments, for example) there *could* be a type of conceivability on which the conceivability of S is knowable a priori, on which S is conceivable simply in virtue of the content of the concepts alone.

As the notion of conceivability is actually *employed* in defending the conceivability of zombies, there is no reason to believe that conceivability is ever *not* relative to an epistemic agent, and every reason to believe that- whether the kind of conceivability at issue is “prima facie conceivability” or “ideal conceivability”- the conceiver stays in the picture.

5.1a Peering Into the Possible?

5.1a1 Yablo’s criticism- Any conceiver could be mistaken.

Positive conceivability involves imagining a certain configurations of objects and properties and deciding whether that “modal imagining” verifies a particular scenario, making a considered judgment whether all the details of that scenario could be “filled in” without any contradiction revealing itself. But any ordinary epistemic agent can be ignorant of some fact relevant to the scenario she believes she can conceive. Consider Yablo’s objection to the immediate inference from conceivability to metaphysical possibility. He writes,

If ignorance of an individual's essential properties can generate modal error, why not ignorance of a property's essential properties? Imagine that my grasp of a property S fails to reflect the fact that it is essentially uninstantiable (S might be the property of being sodium-free salt). Nothing to prevent me, then, from conceiving it as possible that Ss should exist: a de dicto conceivability error rather than a de re one. Likewise the de dicto impossibility that some Qs are Rs will be conceivable, if my understanding of Q omits its essential property of having no Rs in its extension. Probably there is no proposition for which a worry like this cannot be raised. (Yablo, 2003:17)

Nothing is to prevent me from *believing* I can “modally imagine” a scenario if I am ignorant of some essential property of a property or concept involved. I cannot prevent myself from making such a mistake, and no community of likewise-ignorant epistemic agents could correct my mistake. Chalmers accepts this point when he writes,

Indeed, it is arguable that one can modally imagine S when S involves an a priori contradiction. An example may be a case in which one imagines a geometric object with contradictory properties. In cases like this, one imagines a situation in something less than full detail. (2002)

Believing that you can ideally modally imagine Penrose’s Triangle would be an instance of just this sort. Our cognitive limitations bar us from knowing whether we have *ideally positively conceived* of a scenario that verifies S or merely *believe* we have positively conceived of such a scenario, and believe that our positive conception passes the tests for ideal conceivability. But our limitations also bar us from knowing whether we have imagined a situation in full detail or we have left pertinent details unspecified, letting a priori contradictions go unnoticed. Yablo rightly warns that there is probably no scenario for which such worries cannot be raised.

This is why positive conceivability must be idealized. But the question remains as to how *ideal conceivability* ought to be explained.

5.2 Idealizing Conceivability?

So far the notion of “ideal conceivability” has been left undefined, but defending its status as a rational notion (like a priority and entailment) depends on there being some account of idealizing conceivability that “abstracts away” from the subjective judgments of a limited (fallible) epistemic agent. At the same time, it be of no use to idealize “conceivability” by appealing to an *omniscient* conceiver, since, by hypothesis, whatever an omniscient agent only *believes* is *true*. Omniscient agents do not “reason” or “make inferences”, they do not “*arrive* at conclusions”, they do not need to perform such epistemic acrobatics, taking careful, rational steps towards true belief; they are, so to speak, (always) *already there*. Introducing an omniscient conceiver runs the risk of reducing “ideal conceivability” to a triviality; the claim that an omniscient conceiver would find scenario S conceivable is tantamount to simply *stipulating* the coherence of the concepts that compose the scenario, granting its the conceivability by fiat rather than by argument.

There are (at least) two ways to explain the notion of ideal conceivability: one would be to appeal to an ideal (though not omniscient) conceiver- an ideally reasonable epistemic agent. Chalmers initially entertains this option,

One could try to define ideal conceivability in terms of the capacities of an ideal reasoner — a reasoner free of all contingent cognitive limitations. Using this notion, we could say that *S* is *ideally conceivable* if an ideal reasoner would find it to pass the relevant tests (if an ideal reasoner could not rule out the hypothesis expressed by S a priori, for example). A strategy like this is taken by Menzies (1998). (2002: 148, parenthetical remark original, emphasis mine)

His explicit worry with this first account is that an ideal reasoner may be impossible- the notion of an “ideal reasoner” may even be incoherent. He speculates, “...it may be that for every possible reasoner, there is a more sophisticated possible reasoner,” (ibid: 148) and

finds this potential problem with the “ideal reasoner” notion to be sufficient justification for moving on to other explanations of how one could idealize conceivability.

The second potential definition of ideal reasoning, or ideal conceiving, the one upon which he settles, does not employ the judgment or capacities of any epistemic agent but appeals to the notions of justification and defeasibility, properties that do not seem to be reasoner-dependent. He writes,

Alternatively, one can dispense with the notion of an ideal reasoner, and simply invoke the notion of undefeatability by better reasoning. Given this notion, we can say that *S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeatable by better reasoning.* (2002: 148)

This alternative explanation might avoid the potential “regress of ‘sophisticated reasoners’,” the purported reason for dispensing with the notion of an ideal reasoner. Ideal conceivability relies on a reasoner only insofar as there must be some epistemic agent who makes the judgment that S is *prima facie conceivable*.

The idea is that when prima facie conceivability falls short of ideal conceivability, then the *claim* that the relevant tests are passed will either be *unjustified*, or the justification will be defeatable by further reasoning. For ideal conceivability, one needs *justification that cannot be rationally defeated*. (ibid: 148, italics)

There are several problems with Chalmers’s rejection of the “ideal reasoner” explanation of ideal conceivability. The most pedestrian is that the account he endorses in its stead, invoking the notions of “undefeatability” and “justification”, seems open to the same regress problem for which “the reasoner” was dismissed. The strength of justification for a particular claim (i.e., whether it is “defeatable by further reasoning” or “cannot be rationally defeated”) is determined by several factors, one of which is the means by which the subject acquires the belief, opening up the possibility of a regress of *‘better means of acquisition’*,

analogous to the regress of '*the more sophisticated reasoner*',¹¹² the justification for dismissing the "ideal reasoner" explanation. For any given means by which a subject acquires a belief, it is possible (or at least conceivable) for there to be some superior means by which that belief could be acquired, a means that would create a stronger rational connection between the believer and the object of her belief, conferring stronger justification on that claim or belief. There is much more to be said about justification, especially justification of one's beliefs about their immediate phenomenal states. Here I am only pointing to the possibility of an analogous regress.¹¹³

The larger problem with "dispensing with the reasoner" is that, conceptually, *the reasoner is not something with which we can dispense*. Defining "ideal conceivability" in terms of "undefeatable justification" rather than what "an ideal reasoner could not rule out," will not remove the *conceiving subject* from conceivability.

¹¹² Since the possibility of a "perfect reasoner" (as opposed to a merely "highly sophisticated" reasoner) reasoner was left aside when considering the regress-of-sophisticated-reasoners, I will leave aside the possibility of divine belief-inception here, which is tantamount to the "omniscient conceiver" possibility dismissed above.

¹¹³ One might object that there is, in theory, an ultimate, non-miraculous means of belief-acquisition: *acquaintance*, similar to Russelian acquaintance with sense data. Subjects may be related to the object of their phenomenal beliefs (the phenomenal properties of their conscious experience). The dualist may be counting on acquaintance as the regress-of-better-justification stopper, as I believe he is, and I believe Chalmers uses it to argue for the epistemic inequality of our phenomenal beliefs and whatever ersatz "phenomenal" beliefs our zombie-twins hold.

In (Ross, (2013-in progress)) I examine the dualistic acquaintance relation and the epistemic connection between subjects and the objects of their phenomenal beliefs. Surprisingly, even if the property dualism is true, and we are acquainted with the properties of our conscious experience in the way they claim, and our epistemic relation to the objects of our phenomenal beliefs is more intimate than any epistemic relation we could bear to a physical property, acquaintance is in fact a less secure epistemic relation than we usually assume, and the phenomenal knowledge acquired via acquaintance *less certain* than we are usually led to believe. Unfortunately this analysis is outside the scope of our current discussion.

5.2a The (Less-Than-Perfectly) Rational Notion of “Ideal Conceivability”

If “ideal conceivability” is a rational notion, it stands apart from the exemplary rational notions of “a priori” and “entailment”. “Being true a priori” and “being entailed by” are properties of propositions (or sentences) *themselves*; in the first case, the property can belong to an individual proposition; in the second case, the property belongs to a set of propositions (i.e., proposition P entails proposition Q). To be a “rational notion”, it seems that ideal conceivability should be a property of propositions themselves, as Chalmers claims it is.¹¹⁴ But unlike the *purely* rational notions, conceivability- even ideal conceivability- is not a property of statements or propositions, or set of propositions, *in themselves*. The property “being justified” belongs to a *belief*- or, following the quote above, we could say it is a property that belongs to claims and judgments. A belief by its nature not merely a proposition but a kind of propositional attitude, and as such requires a *believer*, a subject who forms and holds that belief and in whose cognitive economy that belief can be instantiated.

Up to this point we have been talking of beliefs as mental representations,¹¹⁵ or perhaps propositions, and this characterization has been sufficient for our purposes. But insofar as beliefs are the kind of things that can be *justified*, beliefs are not merely propositions. ‘Being justified’ is a property of a proposition only insofar as that proposition is a belief. And the kind of justification that belongs to a particular belief, or the extent to which

¹¹⁴ See Chalmers (2002).

¹¹⁵ As I mentioned earlier, this discussion assumes a mental representation account of belief, the same the account that Chalmers’s uses in his theory of the epistemology of phenomenal belief. On this view, beliefs are something akin to sentences in a Language of Thought, and concepts are treated as units out of which beliefs are composed, roughly in the manner that words compose sentences. This way of characterizing belief does not necessarily represent the views or opinions of the author.

that belief is justified, depends on the kind of epistemic relation that holds between a believer and the object of her belief. The *believer* is not something with which we can dispense without losing precisely that entity to which the property ‘being justified’ belongs: the *belief* itself.¹¹⁶

Acts of *conceiving* and *reasoning*- ideal or otherwise- are epistemic acts. As such, the notion of an *epistemic agent* or subject is integrated into the very concepts of *conceiving* and *reasoning*. *Conceiving*, *reasoning*, and *modally imagining* require some subject who possesses concepts and can make inferences. Thus a coherent notion of “ideal conceivability” will be one in which there is a conceiver, and it will be “ideal” insofar as that conceiver has a set of specified idealized properties.

5.3 IREA: The Ideally Reasonable Epistemic Agent

There is a conceptually coherent notion of “ideal reasoner” or “ideally reasonable epistemic agent” in the vicinity, one that conforms to the idea that the conceivability of a scenario is determined by the content of the concepts composing the scenario, and can accommodate Chalmers’s claim that to “positively conceive” of a scenario is to “modally imagine” an arrangement of objects and properties. This would be the IREA- the Ideally Reasonable Epistemic Agent. She is an *ideal* epistemic agent in that she makes judgments about the relations between the contents of her concepts that are *as flawless* as the judgments she can make about the content of her concepts themselves. Whether or not a scenario is

¹¹⁶ We might even say that a belief is a kind of *relation* (if being a type of propositional attitude is being a type of relation), and the relata in the belief-relation are *epistemic agents*, on the one hand, and *the objects of an epistemic agent’s intentional mental states*, on the other.

actually conceivable is entirely determined by the content of the concepts involved in the scenario being evaluated, so any scenario the IREA can "modally imagine" without encountering any conceptual incoherence or conflicts should be ideally positively conceivable and, according to CP, we can infer its metaphysically possibility. As an IREA is ideally reasonable but not an *omniscient* epistemic agent, she makes conceivability judgments based on the content of the concepts at her disposal insofar as she has epistemic access to that content.

All actual epistemic agents are limited in at least two respects- their cognitive capability and the scope of their empirical knowledge. We can "idealize away" from a subject's limited empirical knowledge- as we do when we conceive of Mary as "knowing all the physical facts" or "being scientifically omniscient"- without making IREA herself omniscient. But the judgment of any conceiver, whether an IREA or an ordinary epistemic agent, will always be limited by the epistemic access she has to the content of the concepts composing a scenario, that is, to the content of her own concepts. An IREA is not in principle *infallible*, since she is not stipulated to be omniscient, but her judgments must be trustworthy or conceivability judgments would be useless in philosophical arguments.¹¹⁷

All concept users make judgments based on their beliefs about the content of their concepts, and the accuracy of these judgments depends- in part- on the epistemic access she has to this content of the relevant concepts. It seems that we have some sort of privileged

¹¹⁷ Insofar as a concept has determinate content there will be one determinate and objectively accessible answer to questions regarding conceivability- yes or no. Imaginability is a different issue- x may be imaginable by S but not by S*, as young children can "imagine" collapsing a three sided plane figure without altering the length of any of the sides. But conceivability- as it plays a role in conceivability arguments- is decided solely on the grounds of conceptual consistency, which is determined by the content of those concepts. This has been our working assumption since we began and ought to be indisputable.

epistemic access to the content of our concepts, especially our phenomenal concepts. Perhaps our access is not perfect- we cannot assume that we have *transparent* access to the contents of our minds- but our access must be good enough for us to have very high confidence in the accuracy of our conceivability-judgments. Without sufficient access to the content of the concepts that compose a scenario S we could not infer the possibility of S from our judgments of its conceivability alone.

But even if ordinary epistemic agents did have perfect access to the *content* of their concepts, they are still fallible and could make mistakes regarding the *relations between* the content of these concepts. I may know P, and- at least implicitly- I may know that P implies Q, but I may not conclude that Q due to some error in my reasoning.¹¹⁸ When we conceive of an IREA, we can “abstract away” from this- genuinely contingent- cognitive limitation.

IREA is not burdened with *this* epistemic weakness; she makes *the most reasonable judgments possible given the evidence at her disposal*. Ordinary epistemic agents have the same evidence at *their* disposal, so the difference between and ordinary epistemic agent and an ideally reasonable epistemic agent is not difference in *kind*; both epistemic agents gather information from the world in through the same means, and perform the same kind of epistemic activities. The IREA merely takes this information and draws the most reasonable conclusions from it.

The difference between the two is actually a matter of *degree of reasonableness*; an IREA never falters when contemplating the contents of her concepts to the *full extent* of her

¹¹⁸ Perhaps the chain of reasoning from P to Q is extraordinarily long, and though I know all the “links” in the chain, its length gives me ample opportunity to make a mistake somewhere- a “wrong turn” along the path that *should* lead from P to Q, so I conclude $\sim Q$.

cognitive capacity, while ordinary epistemic agents can easily fail in this respect. Our *epistemic access* to the *content* of our concepts should be on a par with that of an IREA; otherwise we could not begin to guess what an IREA could or could not modally imagine. For the notion of an IREA to be a useful tool in evaluating conceivability judgments, an IREA must judge the conceivability of any scenario based on the same concepts with the same content as you and I would judge them. Returning to the notion of ‘justification’, an IREA’s conceivability judgments will be the most justified, undefeatable by better *reasoning*, since she unflinchingly contemplates the contents of her concepts to the *full extent* of her cognitive capacity. Her powers of contemplation and concentration are idealized, and she is *ideally reasonable* regarding the connections between the content of her concepts to which she has epistemic access.

For the purposes of the upcoming thought-experiment, we do not need to answer the question of *how* we should discriminate which creatures have concepts and which do not. This can remain open, so long as we acknowledge that any creature that has language certainly has concepts, and grant that any creature that has a *symbol system functionally equivalent to a language* will have something like “syntactic units” which are *functionally equivalent* to “real” concepts. Since we have language, we have concepts. And since language use is part of our behavior, our zombie-twins will have something- a symbol system- functionally equivalent to human language. If we assume that zombie worlds are conceivable, then IREZAs are conceivable as well: Ideally Reasonable Epistemic Zombie-Agents.

5.4 IREZA: An Ideally Reasonable Epistemic Zombie-Agent

As zombies have concepts in the relevant sense, zombies can make judgments based on the content of their concepts insofar as they have epistemic access to that content. Regardless of whether the “terms” in the zombie-language are *meaningful*, the symbol system will be functionally identical to our language, so the “terms” will be contentful in a functional sense. Making judgments or inferences about conceivability is a functional process, and the accuracy of such judgments is an objective matter, so, for the purposes of *reasoning*, a symbol-system functionally identical to a meaningful language is just as good as a meaningful language itself. Zombies will make inferences and conceivability judgments in a manner functionally identical to their human-twins. Judgments about conceivability are not “subjective” in the sense that the accuracy of the judgment depends on *the character of the experience* of having a belief; a subject's particular subjective experience, or lack thereof, is irrelevant to their capacity to make conceivability judgments. Genuine conceivability, or ideal conceivability- the only kind that is relevant to conceivability arguments- is an objective matter.

As IREZA is IREA’s zombie-twin, her functional duplicate, IREZA’s judgments will be functionally equivalent to IREA’s; they will look, when written, and sound, when spoken, identical to IREA’s expressions of her judgments. An IREZA may be currently pressing keys in the same manner as an IREA, leaving marks on a screen that, as expressed by *IREA*, are about conceivability judgments. Her zombie-twin may spent many years reasoning about conceivability and the relations between concepts, though when IREZA “reasons about the relation between phenomenal concepts” she is reasoning about the relations between *zombie-*

“phenomenal” concepts. And while the expressions that represent the scenarios they contemplate are phonetically identical, the scenarios about which IREA makes conceivability judgments are composed of phenomenal concepts, and the scenarios about which IREZA makes her conceivability-judgments are composed of zombie-“phenomenal” concepts.

However the content of a zombie-“phenomenal” concept is ultimately defined, one necessary feature will be that the concept does not have phenomenal content. So while IREA’s and IREZA’s judgments are *functionally identical*, they are making claims about the conceivability of *different scenarios*, since the scenarios they contemplate are composed of *different concepts*.

5.4a Experts without Expertise

An IREZA's judgments about the *relations* between the contents of her concepts are *as flawless as* the judgments she makes about the *content* of her concepts themselves; they must be, if IREZA is IREA’s functionally identical zombie-twin. But the IREZA is a *zombie*, and as such there will be problems with her judgments involving zombie-“phenomenal” concepts. Though she is ideally reasonable, the IREZA cannot *know*, she cannot make the *judgment*, that her zombie-phenomenal concept has *no phenomenal content* (at least in part because the IREZA has no knowledge of phenomenality, as she is not acquainted with any phenomenal states).

The IREZA is “without a certain perception”- she has no epistemic access to the non-existence of phenomenal content within her concept.

Just as the dualist and a posteriori physicalist claim that an ideally reasonable epistemic agent would judge zombies and inverters to be conceivable, IREZA will judge that

the content of her zombie-“phenomenal” concepts and the relations between them allow her to conceive of subjects, individuals *qualitatively identical herself*, whose physical and functional features are identical to each other but whose “phenomenal” experiences are the inverse of one another. But just as the schmenomenal-zombie made mistakes in her conceivability judgments, *the IREZA* will be wrong as well; though she is more reasonable than an ordinary zombie, her epistemic access to the content of her concepts is no more transparent- she is no more “enlightened” regarding the concepts of her zombie-”phenomenal” concepts- than her fallible zombie-kin.

5.5 Conceptual Incompetence or Hallucination? The Undetectable Errors of an Ideally Reasonable Epistemic Agent

The IREZA’s conceivability judgment is intelligible to *her*; to try to make her judgment intelligible *to us* we might describe her reasoning process in this way: when the zombie conceives of an invert she is conceiving of a subject Z (just like (or qualitatively identical to) herself) whose “qualia” (the zombie-pronunciation of the referent of the zombie-phenomenal concept) are inverted relative to the “qualia” of the Z*, though the physical, functional, and behavioral properties of Z and Z* are identical. But the IREZA *cannot* be conceiving of what she *takes* herself to be conceiving, because *the content of her zombie-“phenomenal” concepts does not allow her to conceive it*.

When the IREZA contemplates “phenomenal inversion”, she is thinking about “*phenomenal*” *inversion*, whatever that may be. The only beings that IREZA could be conceiving are those whose beliefs or states have “phenomenal” content. the inconceivable

“Inverted-Zombie-Twins” and “Schmenomenal-Zombie-Zombie”. These are the names of conceptually incoherent notions; the schmenomenal-IREZA is not actually conceiving of creatures that *could* instantiate the *incompatible properties* these notions represent. “Inverted Schmenomenal-Zombie-twins” and “Schmenomenal-Zombie-Zombies” are logically impossible and hence by our criteria *inconceivable*, though this does not stop us (or the schmenomenal-IREZA) from finding these *names*- the linguistic expressions representing the inconceivable scenarios- to be *intelligible*.

Even if we treat zombie-phenomenal concepts as empty, as having no referent in the zombie-world (as the property dualist would prefer), (empty-Zombie-)Inverts and (empty-Zombie-)Zombies will still be conceptually incoherent. If there is no referent for zombie-phenomenal concepts that takes the place of “the phenomenal” in the zombie world, then (empty-zombie)-Zombies will be conceptually incoherent, as there will be *nothing* that could be *removed* from a zombie-subject while leaving all her physical, functional, intentional, structural, etc., properties just as they are. If IREZA’s zombie-“phenomenal” concepts are empty, they will not allow her to conceive of “phenomenally” Invert-Twins. When the IREZA takes herself to be conceiving of an invert, the emptiness of her zombie-“phenomenal” concepts would make it that, *if* she were to be conceiving of anything at all, she would be conceiving of a subject with *nothing* being inverted as compared to *nothing*. And the notion that ‘*nothing* could be "inverted" relative to *nothing*’ is conceptually incoherent.¹¹⁹ Of course, the zombie finds the description of the scenario *intelligible*, but, as

¹¹⁹ By the same token, (zombie)-zombies will be conceptually incoherent. Zombies cannot conceive of zombies; if the zombie-phenomenal concept has no content, there will be *nothing* to conceive of as being *taken away* from the *nothing* that is already the zombie's state. If the zombie-phenomenal concept is a schmenomenal-concept, the schmenomenal-zombie will be in the same position as she was when she “took herself to be

usual, we cannot take this *sense of intelligibility* to show that she is actually conceiving of an inconceivable scenario.¹²⁰

The IREZA mistakes an *intelligible expression* for some *conceivable scenario*, a scenario which (since it is composed by zombie-“phenomenal” concepts”) is in fact conceptually incoherent, and she is *constitutionally incapable* of detecting this error.

How do we make sense of the “spotty” epistemic access an IREZA has to the content of her zombie-phenomenal concept? We could say that she mistakes the scenario she is actually evaluating (Zombie-Invert-scenario) for the scenario we would be evaluating if the scenario were composed of actual phenomenal concepts (Invert-scenario). But this will not do, because the zombie has no concepts available to her with which she can make this sort of “mistake”; she can neither “take” nor “mistake” herself as having a thought, J, about object S *when it is conceptually incoherent (inconceivable, logically impossible) for S to be the object of a zombie-propositional attitude, thought S*. We might say she is blind to the non-existence of phenomenal content in her schmenomenal-concept, or to the absence of content in her zombie-phenomenal concept, but this doesn’t quite describe the case, as one cannot be blind to something that does not exist. We *can* be blind to something’s *absence*- for example (to manipulate Sartre’s scenario of Pierre and the café) I can either see that Pierre is absent from the café or fail to notice his absence. But if Pierre does not exist, has never existed- *if I have*

conceiving” of a (zombie)-Invert By this line of reasoning, conscious human beings will be conceptually incoherent as well, though every IREZA- every zombie- will judge humans to be conceivable (and him or herself to be a human).

¹²⁰ Nor should we take the zombie finding this expression *intelligibility* to imply that she is conceiving *what that expression would represent* were the scenario to be conceptually coherent, e.g., a scenario represented by a phonetically identical expression but be composed of *phenomenal* concepts rather than *empty zombie-phenomenal* concepts.

no concept that has Pierre as part of its content or its referent- I cannot fail to notice his absence. He is not “absent”; he does not exist. There is no “he”, at least not for me. And there is no “phenomenality”, at least not for my zombie-twin. Pierre’s non-existence is not the kind of “fact” that could be epistemically available to me- and the non-existence of phenomenality is not the kind of “fact” that could be epistemically available to my zombie-twin- these are not “facts” or “things” or conceptual content that my zombie-twin and I can either notice or fail to notice.

We could say that she is conceptually incompetent in wielding her zombie-phenomenal concepts, but this is awkward as well, since she *is* ideally reasonable concerning *all the content to which she has epistemic access*, which is maximal competency for an epistemic agent. Put in these terms, an IREZA is *just as competent* as an IREA.

We may or may not want to say that the IREZA holds a *false* belief- if “judging incorrectly” requires the ability to “judge correctly”, the IREZA is doing neither. If we can assign the IREZA’s propositional attitude enough content to call it a *belief* at all, then IREZA holds a conceptually incoherent belief. Possibly, the most sensible interpretation is that IREZA, and all our zombie-twins, are in the grip of a *conceptual hallucination*; an undetectable, inescapable, prison of their own “minds”.

5.5a How do you solve a problem like IREZA?

IREZAs are no mere oddity, no benign curiosity for dualists’ (soon to be growing) Curio Cabinet of Property Dualism. IREZAs are as unavoidably problematic as the IREZA’s own inability to correctly judge the conceivability of zombies. Our only way of evaluating the conceivability arguments for property dualism- our only epistemic access to the

possibility of irreducibly immaterial phenomenal properties- is through our judgments about what is conceivable. If property dualism is true it follows that zombies are not only conceivable but metaphysically possible, so it is metaphysically possible for ideal reasoners to draw inaccurate conclusions from undetectable conceptual hallucinations. If it is possible for ideal reasoners to make undetectable errors due to undetectable but unavoidable cognitive limitations, then “ideal primary positive conceivability” does not *entail* metaphysical possibility; there must be some further criteria such that the conceivability/possibility principle can hold.

The metaphysical possibility of zombies seems to undermine the credibility of conceivability judgments, the judgments by which we determine that property dualism is true. How do we circumvent this property dualist “sink-hole”?

5.6 Conceivability by Fiat

If there is a metaphysically possible world in which Ideally Reasonable Epistemic Agents can make mistakes about what is Ideally Positively Conceivable- mistakes that are both *inevitable* and *in principle undetectable*- then appeal to the judgments of an ideally rational epistemic agent *alone* is not sufficient for drawing conclusions about metaphysical possibility, even though these judgments are “undefeatable by *better reasoning*” (Chalmers 2002: 148, emphasis mine). We must build in further restrictions, and these are our leading options:

Option 1: We must making sure that the IREA who is judging whether a scenario is Ideally Positively Conceivable- who is doing the "modal imagining"- is one who *has transparent epistemic access to all the content of her relevant concepts*.

Option 2: We must stipulate that physicalism is *not ideally positively conceivable*. In order to guarantee that an Ideally Reasonable Epistemic Agent's judgments about the conceivability of zombies imply that zombies are metaphysically possible, we must stipulate that physicalism is false of our world.

But these "options" actually collapse: choosing option 2, and declaring that physicalism is not ideally conceivable, is a way of embracing option 1, attempting to stipulate that IREA's epistemic access to the content of her concepts is unrestricted. If consciousness is an essential property of human beings, then (on Chalmers's view) it is *conceptually necessary* that a human IREA be aware of the phenomenal content of her concepts and beliefs. As Chalmers has said,

"There is not even a conceptual possibility that a subject could have a red experience... without having any epistemic contact with it: to have the experience is to be related to it in this way" (1996: 197).

One way to guarantee that IREA is not "blind" to the content of the concepts that constitute the zombie-scenario and invert-scenario- her phenomenal concepts- is to stipulate that IREA is not a zombie but rather a conscious human being who has intimate epistemic contact with her phenomenal experiences, giving her epistemic access to the phenomenal content of their beliefs and concepts.

This will sound strange at first, but at the end of the day, the IREA/IREZA problem will leave Chalmers in desperate philosophical straits. The only way for him to ensure that all possible IREAs are conscious beings who have unrestricted epistemic access to the content of their experiences, their concepts, and their phenomenal beliefs- thus safeguarding the conceivability/possibility principle- *is to stipulate that physicalism is false*. In the section that follows I will introduce the "anti-zombie" argument (Frankish, 2007), the argument that uses

the conceivability argument itself to push Chalmers into this argumentatively unattractive corner.

5.6a A Double Standard for Intuition? Denying the Conceivability of Physicalism

5.6a1 Chalmers's Response to the "Anti-Zombie" Argument

At root, the conceivability argument for property dualism relies on an intuition- the intuition that zombies and non-physical mental properties are conceivable, or, as Chalmers would put the point, the conceivability argument relies on philosophers making the *prima facie judgment* that the zombie-scenario is conceivable. This is the ultimate foundation for the claim that property dualism itself is conceivable, the starting point for the argument that zombies and property dualism are conceivable in the "ideal" or "rational" sense. Plenty of sophisticated argument follows this foundational intuition, but without the intuition that mental properties might be non-physical, the intuition that property dualism is true, arguments for property dualism would have nowhere to begin, no philosophical ground to stand on.

Frankish (2007) presents an "anti-zombie" argument that brilliantly exploits the dangers of leaning too heavily on intuition. It is intuitively conceivable that property dualism is true, and the property dualist builds on this intuition to argue that zombies, and property dualism, is ideally conceivable. And if the conceivability/possibility principle holds, it follows that zombies are possible and property dualism is true. But if it is also intuitively conceivable that physicalism is true. At the very least, physicalists find it intuitively plausible that physicalism is true, that molecule-for-molecule creatures identical to ourselves would be

conscious in virtue of the physical facts about those creatures and their worlds alone. And if it is intuitively plausible that physicalism is true, then it seems that the truth of physicalism is *as conceivable as* the truth of property dualism.

The same faculty of intuition that supports the conceivability of zombies also supports the conceivability of “anti-zombies”- molecule for molecule duplicates of human beings who are made conscious by physical facts alone. The property dualist claims that the zombie, or the zombie-scenario, is ideally positively conceivable. If an ideally reasonable epistemic agent were to “modally imagine” the zombie-scenario, she could fill in all the details of that scenario without encountering a contradiction. There seems to be no reason not to extend that same charity to the physicalist. If an ideally reasonable epistemic agent were to “modally imagine” the physicalist-scenario, she could fill in all the details of that scenario without encountering a contradiction.

Assuming that the conceivability/possibility principle holds, the conceivability of anti-zombies entails their possibility; that is, it would be possible for creatures physically identical to ourselves to be made conscious by those physical facts alone. If physicalism and property dualism are both intuitively plausible, and if CP holds, then both zombies and anti-zombies are possible. Intuition has now lured us into a contradiction: if physical facts alone make anti-zombies conscious, and the same physical facts hold in the zombie world, then the physical facts of the zombie-world would make the *zombie* conscious.¹²¹ And the concept “*conscious zombie*” is a clearly internally inconsistent.

¹²¹ See also Marton (1998) and Sturgeon (2000: 114–116), Brown (2010) makes a similar argument: if anti-zombies are conceivable, zombies are inconceivable. One of the two (zombie or anti-zombie) must be only *prima facie* conceivable if the other is genuinely or ideally conceivable.

To avoid this undesirable conclusion -if it is undesirable- we must either (a) reject the conceivability/possibility thesis, (b) deny the conceivability of physicalism, or (c) challenge the pronouncements of our intuition. Most physicalists take path (a), denying that conceivability implies possibility; the phenomenal concept strategy does just this. Chalmers takes the third path, (c), denying the conceivability of physicalism itself, at least in any philosophically significant way (2010: 180).

In response to the Frankish “anti-zombie” argument, Chalmers denies that the physicalist’s intuition that physicalism is conceivable- i.e., the physicalist’s *prima facie* judgment that physicalism is conceivable- can support the claim that physicalism is conceivable in the *rational sense* required in order for the conceivability/possibility principle to establish the truth of physicalism.

Chalmers reconstruction of the anti-zombie argument, or what he thinks of as the “meta-modal” argument against the conceivability/possibility principle, is as follows:

- (i) it is at least conceivable that materialism is true about consciousness [it is conceivable that $P \rightarrow Q$] [assumption]
- (ii) it is conceivable that $P \rightarrow Q$ is necessary
- (iii) it is possible that $P \rightarrow Q$ is necessary (by CP, setting aside two-dimensional semantics)
- (iv) $P \rightarrow Q$ is necessary (by CP and S5)
- (v) It is conceivable that $P \rightarrow Q$ is not necessary [assumption]
- (vi) $P \rightarrow Q$ is not necessary

But (iv) and (vi) are contradictory. So one should reject CP. (2010: 180, all text in brackets mine)

The “meta-modal” feature of Chalmers’s reconstruction is beside the point- the anti-zombie argument will go through whether $P \rightarrow Q$ is intended to be necessary or simply true.

For the conceivability argument to succeed, Chalmers requires (i) that the conceivability of property dualism be used as a premise in the conceivability argument and (ii) that CP be a priori, and as such unassailable or irrefutable. But because CP is supposedly a priori, and because the ultimate foundation for the conceivability of property dualism grounded in our intuitions regarding the *content* of the concept of *consciousness*, the dualist must make some furtive claims in his defense of (i) and (ii).¹²² The conceivability argument requires the dualist to use the conceivability of property dualism in his argument for property dualism, but in order to defend CP, the dualist cannot allow CP to imply a contradiction. So while the conceivability of property dualism is a legitimate premise in the conceivability argument, the property dualist must *refuse* to allow the conceivability of physicalism to be used as a premise in a conceivability argument for physicalism.

Hence the “double standards” for intuition.

Chalmers claims that physicalism is *prima facie* negatively conceivable at best, and likely not even *secundi facie*. That is to say, though physicalism seems like it might be a coherent theory of mind at first pass, with a little more rational reflection we will see that physicalism is conceptually incoherent.

It may be *prima facie* negatively conceivable that materialism is true about consciousness, *but [the truth of physicalism] is not obviously conceivable in any stronger sense*. Many people have noted that it is very hard to imagine that consciousness is a physical process. I do not think that this unimaginability is so obvious that it should be used as a premise in an argument against materialism, but

¹²² The conceivability of a scenario- in this case, the zombie-scenario- is determined by the content of the concepts that compose that scenario. If there are substantial disagreements over the content of the essential concepts involved, there will be no way to unambiguously judge the conceivability of that scenario. For the zombie-scenario to be conceivable, “consciousness” must mean “phenomenal feeling” rather than the competing notion of “there being something it “seems like” to the subject to be in this state,” since the second notion of consciousness will find zombies to be conscious no less than ourselves.

likewise, *the imaginability claim [the imaginability of materialism] cannot be used as a premise, [in an argument for materialism, or against CP,] either. Furthermore, if I am right that CP is a priori, then there is an a priori argument that $P \rightarrow Q$ is not necessary, so that it is not even ideally negatively conceivable that $P \rightarrow Q$ is necessary.* (2010: 180, emphasis mine)

To claim that it is not even “ideally negatively conceivable” that physicalism is true (a proper paraphrase of the last line of the quote above), is to claim that physicalism is not conceivable *in any philosophically significant sense*. While property dualism is supposedly ideally conceivable, i.e., genuinely conceivable, physicalism is not. And since “conceivability” and “conceptual coherence” are synonymous in this context, Chalmers is here claiming that physicalism is a conceptually incoherent theory of mind, making physicalism *necessarily false*.

Chalmers must argue physicalism is in fact *inconceivable*. Agnosticism on this point will not suffice, since the truth of property dualism entails the *necessary falsehood* of physicalism. As we noted in the introduction, according to physicalism, every creature in any possible world that is molecule-for-molecule physically identical to a human will be conscious. The mere *possibility* of a zombie-world makes physicalism false- false not only in our world but false in all possible worlds. If zombies are possible, physicalism is *necessarily false*; conversely, if physicalism were *not* necessarily false, zombies would be impossible, from which it would follow that property dualism is an untenable theory of mind.

Denying the genuine or ideal conceivability of physicalism, despite the overwhelming support for its conceivability, as evidenced in part by the intuitions of physicalists themselves, is a contentious move in response to the “anti-zombie” argument or “argument

from parity”.¹²³ If we interpret this claim as “merely” a tacit admission that intuition is not authoritative in conceivability judgments, this would cause serious problems for the dualist as well. The authority of intuition is the bedrock of conceivability, and of conceivability arguments. As we noted in the introduction, this is something Chalmers readily admits,

it certainly seems that [in the zombie-scenario] a coherent situation is described; I can discern no contradiction in the description. *In some ways an assertion of this logical possibility comes down to a brute intuition*, but no more so than with the [mile-high] unicycle. Almost everybody, it seems to me, is capable of conceiving of this possibility. (1996: 96)

There is one type of case in which an assertion of conceivability does not to “come down to a brute intuition”; those in which the content of the concepts composing a scenario are stipulated in advance of our attempt to evaluate the scenario. The a priori truth of proposition P, or the entailment of premise R from Q, only holds within a framework in which the content of the concepts composing P, Q, and R are specified. And many of our concepts do have such “stipulated content”- as I claimed earlier, the “mile-high unicycle” is an example of just such a case.

The content or primary intension of “conscious experience” is supposed to be “phenomenal feeling”, and the evidence for this content is supposed to be our intuition regarding what it is to be conscious, not pure stipulation. But when a scenario involves concepts whose content is hotly contested- such as *consciousness*, *mind*, *belief*, or other contested concepts such as *knowledge* and *goodness*- claims about their content of *will*

¹²³ This move is either a straightforward dismissal of physicalism without argument, that is to say, the dismissal of the physicalists’ *intuitions* without argument, or is a tacit admission that **intuition** is not authoritative in conceivability judgments. If it is the former, then the conceivability argument would not be so much an argument as a bold-faced assertion that dualism is conceivable and physicalism is not. And the physicalist has no reason to take such an assertion seriously.

“come down to brute intuition”. Thus the conceivability of the scenarios composed of such concepts will be rooted in one’s intuition about their content, and the authority of one’s conceivability judgment only as authoritative as their intuition about the content of those concepts.

The conceivability argument requires the property dualist use a double standard when evaluating the legitimacy of intuitions about the conceivability of property dualism and of physicalism. The physicalist has no reason to accept that the dualists’ intuitions regarding the conceivability of property dualism- or the content of the concept *consciousness*- have any better standing than their own intuitions regarding physicalism or alternative interpretations of the content of *consciousness*. Intuition cuts both ways, and, as the anti-zombie arguments shows, if *physicalism* is *as conceivable as* property dualism, the conceivability argument, as well as the conceivability/possibility principle, will be in a precarious position. Without this double standard for intuition, the conceivability argument could collapse in contradiction.

5.6b A Broken Epistemic Arm

There is either a double standard regarding the authority of intuitions about property dualism and physicalism, which I think there is, or there must be an argument for the claim that the intuition that the primary intension of *conscious experience* is “phenomenal feeling” rather than “*seeming* or *believing* that one’s state has phenomenal feeling” in the functional, cognitive sense. In the past, the knowledge argument was thought to settle this debate; it has been called the “epistemic arm” of the argument for property dualism.¹²⁴

¹²⁴ The conceivability argument is taken to be its “metaphysical arm”.

Mary the Color-Scientist supposedly learns something new when she has a conscious experience of color, perhaps gaining phenomenal information that cannot be garnered in virtue of knowing all the physical facts alone. Mary learns *what it's like* to see red when she leaves her black and white room, and in large part the evidence for this claim comes from her reaction to her new color experience, “seeing-red”. Mary is “surprised and delighted” by her first color-experience, her first experience of phenomenal-redness. Though Mary already knew all the objective physical facts there are to know about *that* color, she cannot identify that color on sight; certainly she does not know *this is red* simply on sight, the way we do. She learns “that seeing red is *like this*,” something she could not learn while in her black and white room.

But there is a surprising problem with the epistemic arm of the argument for property dualism. Even if property dualism is true, Mary *will* identify color on sight. *Identifying* color is a behavior- a physical act- and acts such as identification and re-identification part of the “easy problem” of consciousness. The property dualist acknowledges, and embraces, the “Paradox of Phenomenal Judgment”, or the “explanatory irrelevance” of phenomenal consciousness- there will be a sufficient explanation of every subjects’ behavior that nowhere needs to site phenomenal properties or phenomenal knowledge, only cognitive, functional, physical facts. Mary can leave her room and immediately identify the red rose as red- nothing about property dualism prohibits this telling of the “Mary Story”, and the “explanatory irrelevance” of phenomenal knowledge to *behavior* and *cognitive activity* actively encourages it. Though Mary’s identification of red “on first sight” is compatible with property dualism, her ability to do so strips the knowledge argument of any epistemic

evidence it might have provided for the truth of property dualism. The epistemic arm of the argument for property dualism is broken.

6. Mystery or Paradox? “Explanatory Irrelevance” and the Knowledge Argument

6.1 Phenomenal Belief, Phenomenal Judgment, and Explanatory Irrelevance

Chalmers’s account of the relationship between cognitive, psychological, or functional properties, on the one hand, and phenomenal properties, on the other, faces at least one *prima facie* problem which he acknowledges and calls, “The Paradox of Phenomenal Judgment” (1996: 172ff).¹²⁵ Below is a condensed account of the “paradox”.

We have seen that consciousness itself cannot be reductively explained. But phenomenal judgments lie in the domain of psychology and in principle should be reductively explainable by the usual methods of cognitive science. There should be a [fully] physical or functional explanation of why we are disposed to make the claims about consciousness that we do, for instance, and of how we make the *judgments* we do about conscious experience. It then follows that our claims and judgments about consciousness can be explained in terms quite independent of consciousness. More strongly, it seems that consciousness is *explanatorily irrelevant* to our claims and judgments about consciousness. This result I call the *paradox of phenomenal judgment*. (1996: 177, emphasis original)

This sort of “explanatory irrelevance” is not unique to property dualism- there will be a sufficient reductive explanation of how we produce phenomenal judgments that is

¹²⁵ The “paradox” begins with the problem of the apparent “explanatory irrelevance” of phenomenal consciousness in producing phenomenal judgments, and opens up questions regarding (but not limited to) how our phenomenal concepts refer, the content of our phenomenal beliefs, and the relationship between phenomenal judgments and phenomenal properties. But the particular problems of property dualism and phenomenal belief is the subject of other work- though fascinating, we can put them aside here.

compatible with most common theories of mind, dualist or physicalist; an explanation that does not explicitly feature the object of these judgments- the phenomenal properties themselves.¹²⁶

This is no real problem for the physicalist: if physicalism is true, phenomenal concept and physical concepts refer to the same property while (as dualists and a posteriori physicalists agree) the phenomenal and non-phenomenal *concepts* that denote this property are “isolated” from one another. Statements expressed in physical or functional (non-phenomenal) terms cannot be translated into statements expressed in entirely phenomenal terms without loss of meaning, but these terms share their *referents*. If physical and phenomenal concepts are conceptually isolated, their isolation will account for the absence of phenomenal *terms* from sufficient functional explanations of how phenomenal judgments are formed. But because the physicalist holds that phenomenal properties are identical to or supervene upon physical properties, phenomenal properties will be implicitly present in reductive explanations of how we come to make these judgments. Phenomenal properties do enter into the explanation, but *implicitly*, as denoted by physical/functional concepts.

The “paradox of phenomenal judgment”- the explanatory irrelevance of phenomenal properties in explanations of phenomenal judgments- does not “stick” to physicalist theories of mind. Mind/body identity (for identity theorists) or the supervenience of mental states on the physical (for most other physicalists) shows how “phenomenality” is present (though not explicitly *mentioned*) in their reductive explanation. For property dualism, explanatory irrelevance is more than a prima facie problem- it is at least *secunda facie*. Dualists may also

¹²⁶ I am uncertain whether this would be true of substance dualism or panpsychism, and fairly certain it would be false of a Berkeleyan idealism.

provide a fully explanatory reductive account of how phenomenal judgments are formed as well (given that judgments “lie in the domain of psychology,” and psychological properties have reductive explanations). If phenomenal properties are neither identical to nor conceptually bound to non-phenomenal (psychological/functional) properties, the property dualist’s reductive explanation of how we make phenomenal judgments will not involve any mention of the purported object of the judgment- the phenomenal property itself- nor will the property be implicitly present in the explanation (as the referent of a non-phenomenal term). Phenomenal consciousness in no way enters into these sufficient, physically-reductive, explanations of how we make phenomenal judgments that phenomenal consciousness, even if property dualism is true.

Chalmers denies that the Paradox of Phenomenal Judgment is a fatal flaw in property dualism’s epistemic framework; instead, he suggests that we step around the “paradox” by limiting the scope of “judgment” to “what is left of a belief after any associated phenomenal quality is subtracted” (1996: 174). Judgments themselves are only physical events- speech acts, or brain activity- and it is perfectly compatible with property dualism that every physical event has a complete physical explanation. “Verbal reports are behavioral acts, and are therefore susceptible to functional explanation” (1996: 173). The same reductive explanation can suffice for our phenomenal judgments and those of our zombie-twins. Your phenomenal belief, however, is *partially constituted* by its phenomenal referent (so Chalmers claims), this additional phenomenal content distinguishes your phenomenal belief from your zombie twins’ corresponding belief (or “belief”), maintaining the distinction between human phenomenal beliefs and their zombie counterparts.

The problem I will address here is that limiting the scope of “judgment” to include only physically reducible features of mental states does nothing to *resolve* the paradox.¹²⁷ Though property dualists by and large accept that “explanatory irrelevance” is an unavoidable byproduct of their division between cognitive and phenomenal mental states, “biting the bullet” of explanatory irrelevance does come at a cost and will require a trade off. Though not immediately obvious, “explanatory irrelevance” has surprising implications for the knowledge argument: by endorsing “explanatory irrelevance” property dualism actually *guarantees* that Mary be able to identify every discriminable color simply in virtue of the information she is privy to while in her black and white room.

The explanatory irrelevance of phenomenal consciousness certainly does create a paradox: if property dualism is true of our world, Mary will emerge from her black and white room and identify colors *directly on sight*, just as we do. And if physicalism is true, the same strange conclusion will follow. Explanatory irrelevance is the beginning of the end for property dualism. If property dualism is true, Mary should learn something - a non-physical fact- upon seeing red. But property dualism may be true and Mary will still emerge from her room able to identify colors on sight. So Mary’s story- properly told- lends *no support* to property dualism.¹²⁸

¹²⁷ While this stipulation may succeed in distinguishing our phenomenal beliefs from those of our zombie-counterparts’, there is more to the problem of explanatory irrelevance than the comparative state of our zombie-twins’ “phenomenal” beliefs. This quick fix for the problem of zombie-phenomenal belief may cause just as many problems as the original “paradox”. In skirting the issue of zombie-phenomenal beliefs by separating *judgments*- cognitive acts- from full-blooded *beliefs*, the property dualist leaves himself with the challenge of reuniting the act of judging with the content of the belief such that our (real) phenomenal judgments can be expressions of our (real) phenomenal beliefs. (But this is a subject I address in other work)

¹²⁸ Though, in principle, *anything* Mary might *do* upon leaving her black and white room *would be consistent with property dualism*, since property dualism is not falsifiable by empirical evidence.

If Mary will leave the black and white room identifying color on sight *even if* property dualism is true, then the knowledge argument is only a platform for philosophers to examine their intuitions about whether Mary would learn something new *anyway, even if* she could perform this immediate-color-identification “trick”. But in that case the knowledge argument dissolves and we are left with competing intuitions; no longer intuitions regarding what will happen to Mary once she leaves her black and white room but *raw intuitions* regarding whether there is something to learn beyond the “physically reducible” facts, whether there are still *phenomenal facts that elude scientific investigation*.¹²⁹

6.1a The Paradox of Ideal Conceivability

If a case can be made for the conceivability of Mary identifying color on sight, it would not be determinate evidence for the truth of physicalism but it should at least show that physicalism is *as conceivable as* property dualism. And if physicalism is genuinely conceivable (or merely as conceivable as property dualism), then Frankish’s “anti-zombie” argument succeeds, and property dualism is left with a paradox.

If conceivability implies possibility, and both physicalism and property dualism are conceivable, both zombies and anti-zombies are possible. But that is impossible. The conceivability argument collapses in a paradox because, at root, the argument lives and dies by intuition. Even “ideal conceivability” cannot be divorced from the conceiver, and even the

¹²⁹ Mary judges that *this is red* immediately. The remaining question whether there is more to her phenomenal belief than her phenomenal judgment alone. If we asked Mary, how would she respond? It has been long settled that an “enlightened” zombie, a zombie that has been given phenomenal knowledge, would not *judge* that anything extraordinary had happened to her. Yet her introspective knowledge, according to the property dualist, would be *richer* than it would have been “pre-enlightenment”.

ideal conceiver's conceivability-judgments depend on their beliefs about the content of their concepts, which will ground out in intuition. This is not the paradox of phenomenal judgment but the paradox of "ideal conceivability".

6.2 Hypersensitive Mary and The Knowledge Argument

Since I am building a case that zombies are no more than prima facie conceivable, it should come as no surprise that Mary the Color Science will be indicted as well. Not Mary herself, really- I believe that Mary's hard work will pay great dividends (both for her and for zombie-deniers), and that she will know "what it's like" to see red simply in virtue of what she learn in her black and white room. It is really the knowledge argument that falters, or the assumption on which it is based- that Mary does not know all there is to know about color experience before she leaves her black and white room. Frank Jackson, Mary's creator, describes her situation as follows:

Mary is confined to a black-and-white room, is educated through black and-white books and through lectures relayed on black-and white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. (1986: 291)

Mary's scientific color-knowledge is exhaustive; she knows all the scientific facts about color and human color experience that there are to know, that is, all the color-facts one can learn without actually having color experience. In the thirty years following her inception, Mary's story has been extensively revised. Subsequent versions have expanded the ways in which she is prohibited from having color experience; her skin was once painted

shades of grey, she was forbidden from rubbing her eyes so as not to produce colorful pressure phosphenes, she had been prohibited from dreaming in color... Eventually she was simply re-created as colorblind from birth, All these roads lead to the same Super Scientist: Mary is supposedly neurologically identical to a normal human being except that she has *not yet* experienced seeing color.¹³⁰ And she knows all the scientific facts about color and human color experience- perhaps all the scientific facts, full stop (on some accounts, Mary is simply *scientifically omniscient*).¹³¹

Jackson continues,

If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies...

It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning-she will not say, "ho,hum". Hence physicalism is false. (ibid: 291).

The day that Mary leaves her room- or her color-blindness is cured- and steps out into the world of color, upon seeing her first red object she supposedly has a thought, something

¹³⁰ The claim that Mary has never "been in the state" one is in when one has a color experience would cover more territory than the claim, "Mary has never had a color experience", since Mary might be put into *the state one is in when one has a color experience* by some act of magic, or cosmic ray or other bizarre logically possible means.

We should also specify that Mary- while in her black and white room- is never in the state one is in which *represents the knowledge* one acquires by having been in the state one is in when one has a color experience. This stipulation eliminates the "cheating" accusations that are often aimed at thought-experiments designed to show that it is conceivable for Mary to know what it's like to see color in virtue of what she learns *before she has* a color experience. I will comment on this briefly again later.

¹³¹ Mid-Jackson-quote: Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world, W(P), for which it is complete, and the difference must be in non-physical facts: for our world and W(P) agree in all matters physical. Hence, physicalism would be false at our world (though contingently so, for it would be true at W(P)). (1986: 291)

like "oh, so *that's* what it's like to see red!" As Jackson says, she will not say "ho, hum". Her reaction to seeing red is evidence that she learns something upon her first red-experience- she learns *what it's like* to have an experience of red.¹³²

In order to see the anti-materialist argument within the Mary thought-experiment we can look to Lycan's (2003) helpful formalization. In this version of the thought-experiment, Mary is colorblind from birth (the rest of her cognitive faculties intact), and becomes omniscient regarding all (physically reducible) scientific facts. Eventually her condition is cured.

- (1) Before her cure, Mary knows all the scientific and other "objective" facts there are to know about color and color vision and color experience, and every other relevant fact. [Stipulation.]
 - (2) Upon being cured, Mary learns something, viz., she learns what it's like to experience visual redness. [Seems obvious.]
 - ∴ (3) There is a fact, the fact of what it's like to experience visual redness, that Mary knows after her cure but did not know prior to it. [1,2]
 - (4) For any facts: if $F1 = F2$, then anyone who knows $F1$ knows $F2$. [Suppressed; assumes simple factive grammar of "know."]
 - ∴ (5) There is a fact, that of what it's like, that is distinct from every relevant scientific/"objective" fact. [1,3,4]
 - (6) If materialism is true, then every fact about color experience is identical with some physiological, functional, or otherwise scientific/"objective" fact.
-
- (7) Materialism is not true. [5,6]

(4) is supplied because without (4), there seems no way to get (5) from (1) and (3). (Lycan, 2003, 385)

This formulation of the knowledge argument is problematic for anyone who would like to suggest an alternative interpretation of the Mary thought-experiment. Premise (2)

¹³² In 1998, Jackson recants: "The redness of our reds can be deduced in principle from enough about the physical nature of our world despite the manifest appearance to the contrary that the knowledge argument trades on. This is why I now think that the knowledge argument fails." By then the "Mary" bug had bitten philosophy of mind; thousands of articles and oceans of ink have spilled over the captivating super-scientist.

assumes that Mary cannot know what it's like to experience red solely in virtue of what she learns in her black and white room. Certainly, this seems as obvious to most as premise (2) itself.¹³³ But a philosopher who would rather avoid using the locution “knowing *what it's like*” to express the notions of *having knowledge about color experience* has no way of engaging with the knowledge argument in its standard formulation. A priori physicalists would argue that Mary achieves an understanding of red-experience while in her black and white room or prior to her “cure”. But on the standard formulation of the knowledge argument, the a priori physicalist’s claim will be viewed as being accused of begging the question against the knowledge argument by claiming that Mary already knows “what it's like”, to experience red, or has “phenomenal knowledge” of redness, before leaving the black and white room.

In the formalization above, the justification for premise (2) is its apparent obviousness¹³⁴. At the end of the day, I agree that the only support for premise 2 is that it seems obvious to most people, philosophers and normal alike. But there ought to be an intermediate premise (1.5), explicitly noting the empirical evidence for the claim that Mary learns something new when she steps into the colored world.

Premise (1.5) When Mary sees color for the first time, she will not be able to identify colors on sight *solely in virtue of what she learns before her first experience of color*.

¹³³ I know of very few philosophers who disagree with this supposition, who think that Mary will leave her room and not be “surprised or delighted” but rather already know that a red tomato looks like *that*. Those who are compelled by what Beaten (2005) calls the “Mary Intuition”, that,

...Mary, in the circumstances described, will still learn something on first seeing a coloured object (equivalently, that there is something that Mary, in the circumstances described, does not yet know, namely what it is like to see in colour. (4)

¹³⁴ Lycan envisions (the original) Jackson’s response to a philosopher who denies (1.5), a Premise (2) “denier” as, “Yeah, OK, maybe... But it sure does seem that Mary would not know what it's like to see colors without having experienced color in some way” (386.)

Premise (1.5) is an implicit assumption; it is also an assumption that makes premise (2) *seem obvious*. If Mary *could* identify colors (at first sight) in virtue of what she learns before her cure, premise (2) would certainly seem *less* obvious. And the claim that Mary may emerge from her room with the ability to identify colors immediately upon seeing them is not question begging- it implies nothing about whether she does or does not know *what it's like* to see color, or whether she has “phenomenal knowledge” of redness, before she has her first seeing-red experience. The simple suggestion that Mary steps out of her room equipped with this *capacity* does not even touch the issue of “phenomenal knowledge”; the scope of this capacity is limited entirely to her cognitive mental properties. This way of putting the point merely *reconfigures* the question “Will Mary know *what it's like* to see red?” in neutral terms such that the a priori physicalist has the opportunity to propose an alternative interpretation of the Mary thought-experiment without begging the question against the dualist or a posteriori physicalist.

We can see how this non-question-begging “reinterpretation” might be employed by looking at Lycan’s final formulation of premise (4),

(4’’) “If $F1 = F2$, then, barring pronominal discrepancies and so long as $F2$ is effable, anyone who knows $F1$ and is not suffering from scientific ignorance can work out and thereby come to know $F2$ ” (Lycan 2003: 390).

If we interpret this premise in terms of “identifying red” rather than “knowing *what it's like* to see red”, the $F1$ in premise (4) will be facts *on the non-phenomenal side* of the dualist’s ontological divide, i.e., the cognitive, functional, physically reducible facts. $F2$ will not be a fact *per se* but an act or disposition, the *ability* to identify red on sight without acquiring knowledge of facts that the property dualist has relegated to the other side of the divide, phenomenal facts or perhaps phenomenal information.

The essential premise of the knowledge argument is that Mary learns something upon seeing red for the first time, and the evidence supporting this premise is that she will not know the identity of the colors she sees when she leaves her black and white room, at least not *on sight*. If Mary enters the colored world, already equipped with the ability to identify colors on sight, one would need to look elsewhere for evidence that it is logically impossible, or inconceivable, for a subject to know everything there is to know about color experience- or to acquire “phenomenal knowledge”- in virtue of knowing all the physically-reducible facts about color and color experience.

6.2a A Priori Physicalism, and the Significance of “Knowing Red on Sight”

The nearly undisputed assumption is that for Mary to “know colors on sight” she must have a special kind of recognitional concept- vis., a phenomenal concept, such as *red_{ph}*- that she will acquire when she has a subjective conscious experience of redness.¹³⁵ In virtue of knowing all the scientific facts about red, Mary can conceive of red under certain physical and functional concepts, but no depth or breadth of knowledge couched in physical or functional terms could provide Mary with the concept *red_{ph}*; no level of facility with these kinds of concepts will amount to Mary knowing *what its like* to see red. Phenomenal

¹³⁵ In opening his attack on Dennett’s and Mandik’s interpretation of the knowledge argument, Alder (2008) explains that there is no “experience requirement” for acquiring a phenomenal concept, that one does not have to have “an experience with the phenomenal character x” to know what it’s like to have experience x. Referring to Hume’s missing shade of blue, he says “one can extrapolate [knowledge of the phenomenal character of x] from phenomenally similar experiences”, or, citing Lewis (1988: 500), by magic and similar to Unger (1966, 54), by some cosmic coincidence (247). My argument is substantially different from Dennett’s and Mandik’s and makes no assumptions regarding an “experience requirement” on phenomenal knowledge.

concepts are *isolated* from all non-phenomenal concepts, and it is only in possessing phenomenal concepts (such as *red_{ph}*) that a subject “knows color on sight”.

Perhaps it is *prima facie* conceivable that Mary could come know the color of the object in front of her without being *told* its color, but if she can achieve this feat it will only be by taking extreme measures; using a photometer to measure the light-waves reflecting off the colored object,, monitoring her brain activity through a cerebroscope, etc. She can then perform some vastly complex calculation to determine the color of the object. Because Mary does not have the phenomenal concept ‘*red_{ph}*’, she must infer the color of the object, while subjects who have had a conscious experience of seeing red do not *infer* the color of the tomato- they know it directly because they have the special recognitional phenomenal concept *red_{ph}*.

At the other extreme, the a priori physicalist claims that Mary knows everything there is to know about color perception and color experience in virtue of knowing all the relevant physically reducible about color and human color experience, and her ability to use this information at the right time, in the right way. If Mary were to leave her black and white room and identify red as quickly as a seasoned color-seer, the a priori physicalist’s claim seems to be well justified, or at least *plausible*.

Would this show that Mary knows everything there is to know about color and human color experience in virtue of what she learned in her black and white room?¹³⁶ The first step toward answer this question and properly evaluating the Mary thought-experiment is to re-

¹³⁶ What non-question begging principle would allow us to differentiate between (a) *whatever kind of color-knowledge* Mary displays by identifying colors on her first encounter without a moment’s hesitation and (b) “*actually knowing colors on sight*”?

evaluate our conception of Mary's cognitive faculties- just how brilliant is Mary, the brilliant color scientist?

This much seems obvious: if we conceive of Mary's cognitive abilities on the model of an ordinary-though-exceptionally-brilliant epistemic agent, she would not be cognitively equipped to make use of the vast amount of information that constitutes "knowing all the physical facts". This kind of "normalcy" constraint would be a cognitive handicap *relative to* her scientific omniscience. The information at Mary's disposal- "knowledge of all the physical facts"- is not information *for her* unless she can process that information, use it, draw on it at will. There is no point in Mary being omniscient regarding the physical facts of her world unless we also stipulate that Mary can unerringly process this vast information, (recognizing every relevant feature of each physical fact within every possible situation simultaneously and at every moment.)

If "knowing of all the physical facts" is to pull any weight in the knowledge argument it must be this sort of "useful" knowing. Unless Mary's knowledge of all the physical truths includes the ability to *handle* all the physical truths, the knowledge argument gives a priori physicalists no reason to think that Mary's "surprise and delight" reaction to her first color experience is evidence against their position.

Mary's cognitive faculties must be as powerful as her knowledge base is expansive.

6.2b Introducing "Hypersensitivity"

Mary must be aware of the most minute functional fluctuations in her cognitive states, instantaneously identifying the all-but-imperceptible changes in her physiological states and all other non-phenomenal states. Ordinary subjects do not practice or achieve such

heightened self-awareness; this is not the kind of “self awareness” that ordinary people engage in. During her time in the black and white room, Mary is learning facts about color, normal observers, and the entirety of the cognitive or functional states a normal observer would enter into if a normal observer were to see a particular color. And she is learning about these cognitive states and their connection to human color perception down to the minimum distinguishable difference for every humanly distinguishable color. In principle it should be *logically possible* for a subject to achieve this extraordinarily heightened awareness of her own psychological or cognitive states, the states involved in the functional aspects of human activity¹³⁷. Call Mary’s extraordinary awareness of her cognitive states “Hypersensitivity”.

There is no feature of the knowledge argument that would make Mary’s “Hypersensitivity” question begging, nor does it violate any of the argument’s implicit assumptions. And Hypersensitivity is entirely compatible with property dualism. Certainly, no ordinary subject would bother to become Hypersensitive; they would have neither the time nor cognitive faculties to achieve such a feat. But ordinary subjects do not spend their lives in black and white rooms, longing for the day they will finally see color, meanwhile learning every physically reducible fact there is to know about color and human color perception¹³⁸. Ordinary subjects are not equipped with the incredible cognitive capacities Mary must have to learn all the facts she is- by hypothesis- required to learn before she enters the world of color. So it is no surprise that we do not conceive of normal subjects with such a capacity for self-awareness, this heightened degree of awareness of cognitive and functional

¹³⁷ Activities such as identifying color, one among many “easy problems” of consciousness.

¹³⁸ Mary should learn all the facts that supervene upon physical facts as well, since the only knowledge unavailable to her is phenomenal knowledge regarding color. No phenomenal knowledge is actually required for Hypersensitivity; we could replace Mary with Zombie Mary and the point would still go through.

states, to distinguish changes in herself as fine-grained and quickly as Hypersensitivity requires.

Not only is Hypersensitivity compatible with property dualism, property dualism actually suggests, perhaps requires, this interpretation of Mary's story. Every aspect of Hypersensitivity is *entirely cognitive*- the knowledge gained by exercising Hypersensitivity lies on the "non-phenomenal" side of the metaphysical division of mental states. And the "paradox of phenomenal judgment", the "explanatory irrelevance" of phenomenal properties, phenomenal knowledge, and phenomenal consciousness- in short, the irrelevance of phenomenality to sufficient explanations of why a subject behaves as she does- encourages us to conceive of Mary in this way.

6.3 Mary's Story, From the Beginning

The original Mary Story has two well-known endings: one orthodox, one "anti-establishment". If we think of Mary and her competencies in the way that I believe we are not only entitled by required, as Hypersensitive Mary, we will find that the story ends differently than the orthodox interpretation. What this unorthodox ending says about phenomenal knowledge and phenomenal concepts is a separate but related issue.

On the orthodox telling, when Mary has her first phenomenal-red experience and attends to this experience she is "surprised and delighted". Mary did not know that red would look like *this* and, on most tellings, she does not know which color she is seeing. She cannot identify the color, even though by hypothesis she knows all the physical facts there are to know about *that color* from what she learned while in her black and white condition.

This is what we can observe of Mary. Internally, or subjectively, according to the property dualist, Mary is learning a new non-physical fact. According to most physicalists, she is gaining new information about fully physical entities, coming to know an old fact in a new way, or acquiring a new concept with a physical referent. Both dualists and a posteriori physicalists agree that Mary learns something she couldn't know in virtue of knowing all the physical facts she learned before she had her first seeing-red-experience; she learns *what it's like to see red* or 'that seeing red is like *this*', whether this new "knowledge" is non-physical or additional, ultimately physical, information.

The anti-establishment version of the Mary story ends differently. Mary sees color for the first time and says, "ho hum", or its equivalent; however she reacts, it is not with surprise. If experimenters try a "cheap trick", such as painting a banana blue to fool Mary into thinking yellow looks *like that*, she calls them out on their naïveté.¹³⁹ Mary may react with well-deserved pride- learning all the physically reducible facts about color, and acquiring the faculty of Hypersensitivity, is no easy task. She had faith that she would know that *that* was red upon first sight, but since she had never before had a seeing-red-experience she hadn't the opportunity to test her knowledge and skills.

With her extraordinary powers of self-awareness and vast knowledge of human color-perception behavior, it is likely that she could have put herself into that state with the right visual auto-stimulation. But Hypersensitive Mary guards against this temptation, so as not to "cheat", ensuring she does not gain any "phenomenal knowledge" about red while in her

¹³⁹ This alternative ending was originally posed in Dennett (1991), as well as the blue banana thought-experiment.

black and white room. Given the sort of control Hypersensitivity would give someone as cognitively gifted Mary, it is reasonable to think that she could keep herself from going into such a state as easily as she could intentionally put herself in such a state. And given her vast knowledge and Hypersensitivity, if she had accidentally been in a phenomenal-red-experience state she would have known just what state she was in, and the identity of the color-perception she was experiencing.

Hypersensitivity Mary knows all there is to know about color and human color experience by knowing all the physically reducible- and supervenient- facts about her world, having *maximally competent* cognitive faculties with which to draw the proper conclusions from these facts, and her Hypersensitivity.

6.3a Behaving Properly- The Orthodox Zombie Mary Story

There is an orthodox telling of the Zombie Mary story as well, which follows the ending of the orthodox Mary Story. As Mary is omniscient regarding the physical and functional facts of her world, Zombie Mary- her physical, functional, behavioral duplicate- must be omniscient regarding the physical and functional facts of her world as well, including, as always, facts about color perception. And as is required from such a “twin”, Zombie Mary will behave just as Mary does when Mary leaves her black and white room- she will open her door, see red for the first time, and (just as Mary reacts with surprise and delight upon having her first phenomenal-red experience), Zombie Mary’s reaction is the functional, behavioral, and physical equivalent of “surprise and delight”. Both Mary and Zombie Mary exclaim, “Oh, so *that’s* what it's like to see red!”. Zombie Mary utters the phonemes physically identical to Mary’s exclamation, with identical intonation, and the

physical state (brain state) that causes Zombie Mary to produce such sounds is physically identical to the brain state that causes Mary to produce those sounds.

The only difference between Zombie Mary's story and that of Ordinary Mary is that while Ordinary Mary has a phenomenally conscious experience of her visual perception of the color red and in virtue of this gains new phenomenal information, whether it be non-physical or physical-phenomenal information, Zombie Mary has no phenomenally conscious experience of the perceptual event and gains no new phenomenal information.¹⁴⁰ Both women's brains receive identical physical information, process that information in functionally identical ways, and the physical structure of their brains represent identical results of the processing of physical 'red' information.

But- as we are conceiving of *Zombie Mary* as a *Zombie*- *Zombie Mary's* brain stores only the physical and functional visual facts about her "seeing-red" event; it stores no information about any associated phenomenal quality, because *Zombie Mary* has none. *Mary's* brain stores physical and functional visual facts about her "seeing-red" event *and* facts about *what it's like* to see red, or facts about what it's like to be phenomenally conscious of seeing red. Ordinary Mary acquires a new concept with which she can refer to the phenomenally conscious experience "*what it's like* to see red" directly, through the first person perspective, and can think about red in a new way (according to the a posteriori physicalist, she is thinking about an old fact in a new way).

¹⁴⁰ We are only told to *conceive* of *Zombie Mary* as gaining no new phenomenal information- even if phenomenal information is about physical properties such that it would be impossible for any physically-identical *Mary-Twin* to actually fail to gain whatever information *Mary* gains.

McGeer (2003) finds that the Zombie Mary Story undermines the supposed mutually-supportive relationship between the knowledge argument and the zombie hypothesis (the conceivability argument).¹⁴¹

...assuming we have an adequate explanation for why Zombie Mary seems to learn a new fact on leaving the black and white room, it must be agreed that we have an adequate explanation for why Mary herself seems to learn a new fact under the same circumstances. The challenge for anti-physicalists is now to say why Mary's seeming to learn a new fact on seeing the red tomato involves her really learning a new fact, without begging the question the knowledge argument is supposed to establish – namely, that her experienced transformation must consist in her learning such facts. (2003: 310)

Chalmers (2003) gives the proper property dualist response to the orthodox telling of the Zombie Mary story. As she is Mary's functional and physical twin, Zombie Mary shares all of Mary's functional properties, including Mary's outward physical behavior- the "surprise and delight" response. Zombie Mary is neither "surprised" nor "delighted" insofar as those states have *phenomenal* content, but insofar as these responses have *functional* content Zombie Mary's behavior displays all the functional features of surprise and delight. All of Zombie Mary's behavior is due to her functional isomorphism with Mary, her phenomenal conscious twin. This account of Zombie Mary's behavior is consistent with the claim that Mary's behavior is due not only to the physical properties of her brain (and body) but due to the phenomenally properties of that visual experience as well.

The property dualist has bitten this bullet many times before. It is an instance of "explanatory irrelevance". Whether a creature has or lacks phenomenal consciousness is

¹⁴¹ McGeer has the right idea, but if Mary is Hypersensitive Mary, McGeer has the story the wrong way around. Hypersensitive Mary does *not* seem to learn anything, so the challenge for the property dualist is just the opposite: present a case for Mary learning something though she *seems* to know everything there is to know about human color experience. We have an adequate explanation for why Hypersensitive Mary seems *not* to learn a new fact- we need an explanation of why this "*seeming* not to learn anything new" is inconclusive.

irrelevant to genuine sufficient explanations of how that creature makes judgments about phenomenal consciousness. Though “explanatory irrelevance” has always seemed strange, until now it has been treated as a benign piece in the Curio Cabinet of Property Dualism. But I believe the property dualist cannot bit the bullet of “explanatory irrelevance” any longer, that is, not without blowing the head off the knowledge argument.

6.4 The Hypersensitive Mary Conceivability Argument

1. Cognitive states, events, and properties are non-phenomenal properties. [granted by property dualism]
2. All non-phenomenal properties reduce to physical, functional, and structural properties.
3. All facts about non-phenomenal properties reduce to physical, functional, and structural facts. [1, 2]
4. All the facts about cognitive states or processes either are, reduce to, or supervene upon physical facts. [from 1, 2 & 3]
5. For every physical event, including cognitive activity, there will be a fully physical or functional account of that physical event. [from property dualism: “explanatory irrelevance”]
6. Judging that color R is red, or identifying R as red, is a physical or cognitive activity. [granted by property dualism]
7. Mary can acquire Hypersensitivity, where Hypersensitivity is learning to identify one’s own non-phenomenal cognitive states to the finest grain possible (the minimal distinguishable difference between cognitive states and). [assumption]
8. It is logically possible for Mary’s faculties of information processing, at the personal level, to be as fast as any possible form of information processing. [assumption]
9. Mary can identify her own cognitive activity P as cognitive activity P on the first instance of P as quickly as any identification can be made. [from 7 & 8]

10. Mary gains the knowledge that she is in cognitive state P, *where P= identifying R as red*, as quickly as it is logically possible to process information. [from 9]
11. Mary will identify red on sight in virtue of knowing that she is in the cognitive or perceptual state that a normal observer is in when a normal observer perceives a paradigmatically red object.
12. Mary will leave the black and white room with the ability to identify red on sight.

Mary will immediately identify her novel perceptual state as *being the perceptual state a normal observer is in when a normal observer perceives a paradigmatically red object*. When Mary identifies red on sight in virtue of knowing that she is in the cognitive state ‘seeing-red’, should we say that she *making the inference* that ‘*this is red*’, from knowing that she is in the cognitive state that a normal perceiver is in when a normal perceiver sees red? Or does Hypersensitive Mary “*simply*” know that ‘*this is red*’. How should we answer this question? Could we ask Mary? If so, could Mary answer this question?

6.4a Hypersensitive Mary’s ‘Ho, hum’ Rebellion

Hypersensitive Mary is released from her black and white room and encounters a red tomato for the first time. Since all the facts pertaining to an *identifying-red-event* either are, reduce to, or supervene upon physical facts, and since Hypersensitive Mary knows all such facts and is Hypersensitive,, there is nothing about the cognitive event “normal-subject seeing red for the first time” that can surprise her. She may say, with a tone not unlike joy, “Red! Yes, it is lovely. I’ve been waiting a long time to feast my eyes upon color!” Hypersensitive Mary is duly proud of her genuine scientific omniscience and her

hypersensitive self-awareness; there were years of monochromatic learning behind it. And her captors, themselves neither scientifically omniscient nor as mentally adroit as Hypersensitive Mary, are surprised and delighted to have empirically proved that physicalism is true of their world. (Of course, it is impossible to empirically prove that physicalism is true of our world, just as it is impossible to empirically prove that property dualism is false...)

6.4b Hypersensitive Mary in Nida-Rumelin's "Color Room"

There is nothing *surprising* for Hypersensitive Mary about her psychological reaction to red, even though she has never been in this cognitive state before, because *facts about cognitive events and judgments are, ex hypothesi, entirely physical facts*, which she would know before leaving the black and white room. Given Mary's scientific knowledge and Hypersensitivity to her cognitive states, Hypersensitive Mary cannot fail to know that the color she is perceiving is red.

Vierkant (2002) tells a related version of this story bringing Zombie Mary in Nida-Rumelin's (1993) "Color Room",

... a room where all objects are brightly colored, but not in their natural color. If Mary were led into such a room, she would learn about the color phenomenon, but she would not know which color was which. If she would not know, then neither should the behaviorally equivalent Zombie-Mary. This seems very strange, though. Zombie-Mary knows all the facts there are to know about colors in her physical world, so she shouldn't have a problem with [identifying colors at first sight]. (Vierkant, 2002: section 2.2)

Hypersensitive Mary has her zombie-twin as well, Zombie Hypersensitive Mary, and for Zombie Hypersensitive Mary not to know colors at first sight would be more than "very

strange”. Though this version of the Mary story might be *intelligible*, it is not consistent with the property dualist’s position on cognitive states as fully physically reducible, the explanatory irrelevance of phenomenal knowledge, and the features of Hypersensitive Mary (that she knows all the physically reducible or supervenient facts about color and human color experience, that the speed of her information processing at the personal level is as fast as logically possible, and that she has developed Hypersensitivity).

6.5 “Knowing by sight”?

“Recognizing a color on sight” and “knowing by sight”; though two ordinary, seemingly clear notions, it should be no surprise that they are problematically vague in the context of the knowledge argument. I list four potential interpretations listed below, none of which quite seem to capture their meaning.

1) If “Mary does not know color by sight” means only that Mary’s visual system has not processed red stimuli before, therefore her visual system is not processing this color-information in the manner that a visual system typically would *after* it has processed its first color-stimulus, then no, she does not know the color of the blue banana or red tomato by sight. But this cannot be what “by sight” means here, if whether or not Mary “knows color by sight” is supposed to be relevant to the knowledge argument, because this is an undisputed and banal fact about Mary’s visual system and the kind of information it has and has not processed.

2) If “knowing by sight” means “having knowledge acquired from one’s visual system alone”, “knowing by sight” is a useless notion, since not even a seasoned color-perceiver’s visual system performs this function in isolation, and there are no clear

boundaries between the visual system processing color information and a subject coming to know what color she is perceiving.

3) If “knowing color by sight” or means simply “knowing *what it's like* to perceive that color”, where this ‘*what it's like*’ is an instance of the kind of “phenomenal knowledge” that is at the source of the dispute between *a priori* and *a posteriori* physicalists, then “recognizing color on sight” is not theory-neutral and prohibits the *a priori* physicalist from engaging with the Mary though experiment or knowledge argument. If “knowing color on sight” means having the sort of “phenomenal knowledge” that is under dispute, that is *in principle* isolated from physical knowledge, then according to the *a priori* physicalists no one “knows color on sight”.

4) If “knowing color by sight” is supposed to mean “knowing a color immediately”, we then need to distinguish between two notions of “immediate”: the colloquial, “instantaneous”, and the technical philosophical term, “to be unmediated”. The “instantaneous” interpretation of “immediate” is a non-starter. Identifying something on sight (or by any other means) requires visual information processing, and information-processing is not instantaneous. It may happen in the specious present, but the specious present is itself extended in time. This fact about “information processing” applies to both sub-personal processing and personal-level information processing, that is, to undetected and detectable inference and deduction.

6.5a Surprise, or Suppressed Premise?

Perhaps this vagueness in the notion of “knowing by sight” points to an implicit assumption in the knowledge argument, a suppressed premise that is difficult to express

without begging the question in some way. The first option for such a suppressed premise is the “Slow and Laborious” assumption.

6.5a1 The “Slow and Laborious” Assumption

It will be laborious, or at least more time consuming, for Mary to process the information she has at her disposal, vis., knowing all the physical facts about the world, including her own psychological/physiological systems, etc., and in virtue of possessing this knowledge and being affected by a color-stimulus, identify the color as red, than it is for an ordinary person’s sub-personal information processing systems to take the input they receive from the wavelengths of light hitting their retina, etc., and produce a color-recognition output.

The “Slow and Laborious” assumption may seem obviously true. It implies that Mary will identify red more quickly upon subsequent viewings than on her first. Certainly, she will have gained the ability to identify red in a *new way*, and perhaps she gains this ability in virtue of *learning a new fact* about perceiving red (learning “*what it's like*” to experience that color perception) or by acquiring a new concept, special, isolated kind of concept.

I see little reason to accept this premise in the context of the knowledge argument. Given that we have already defined Mary as *omniscient* regarding the physical facts of her world, it would seem ad hoc to then *limit* the ease and speed of her deductive powers (or cognitive faculties). Knowledge of all the physical facts is of no use to Mary (and of no use to the knowledge argument) if we handicap the very faculties with which she can make use of this knowledge. It is not sufficient to grant that Mary can “quickly figure out” what color she is perceiving by accessing her knowledge of all the physical facts of her world. Mary is performing, at the personal level, the functional equivalent of the information processing that occurs for all of us at the sub-personal level when we recognize a color on sight. Of course, the inferences Mary makes in recognizing a color are themselves undergirded by neural information processing, but the amount of information processed does not, by itself,

determine the speed at which the processing occurs. We might find empirical evidence to support this claim, but even this would be inconclusive, since the principles governing the speed and methods of information processing are ever under revision. Regardless, reasoning in this kind of thought-experiment is limited only by what is conceivable; nomological possibility is of no consequence here.

Unless we allow Mary to process information about “worldly” facts, including facts about a normal perceiver’s cognitive states, and her own cognitive or psychological states on the *personal level* as quickly as logically possible, with the speed and ease of the most efficient form of information processing, and to have unfailing, “immediate” awareness of her cognitive states as finely grained as is logically possible, the Mary thought-experiment of the knowledge argument says nothing about the tenability of physicalism, even of a priori physicalism.

Hypersensitive Mary will have no need of cerebroscopes, spectrometers, or other artificial aids to detect changes in her mental states; in fact, in order to identify the rose as red, she needs no immediate knowledge of her current brain state. but only be maximally sensitive to her (by hypothesis, non-phenomenal) cognitive condition. By sensing the minute differences between the psychological effects of different stimuli in different circumstances, Mary (through her hard-won sensitivity-skills) will always identify the color that she is perceiving (in principle, discriminating them to the minimum discriminable difference). Mary will learn everything there is to know about the neuroscience of color, but she may not need to learn the neuroscience in order to identify colors on sight; what Mary really needs is extensive *sensitivity* training.

Alternatively, the suppressed premise may be that to know a color “by sight” is to know it “immediately”, in the technical philosophical sense of “immediate” and “mediated”. If “knowledge by sight” means unmediated knowledge this term must be qualified, since all sensory knowledge will be mediated by various forms of information processing occurring in a host of connected neural systems. More likely “immediate” here means *not mediated by the wrong kind of information*, by knowledge of the wrong kind of facts.

But we cannot *assume* at the outset of the knowledge argument- implicitly or explicitly- that the entirety of information Mary has in virtue of knowing “all the physical facts” is the *wrong kind* of information for knowing what it's like to see red.

6.5a2 The “Wrong Kind of Knowledge” Premise

If Mary, upon leaving her black and white room, identifies colors on first sight, this will show that she already knew all the facts there are to know about perceiving color and will be evidence for physicalism.

Unless her color-identification is mediated by the wrong kind of knowledge, including but not limited to: knowledge not derived from an experience of seeing red (either veridical or hallucinatory), or knowledge of facts about her psychological/physiological systems and dispositions.

In that case her act of identifying-colors-at-first-sight will not be counted as “knowing colors by sight”- it will not show that she already knew all the facts there are to know about perceiving color- and her act will not be evidence for physicalism. She has only inferred that “*this*’ is the sensation of perceiving red” from knowledge that is not itself knowledge of

what it's like to see red; she did not already *know* that "*this*" is the sensation of seeing red; she only had the ability to *infer* it.

If the knowledge argument has such a suppressed premise, then the argument assumes its own conclusion, and it sounds quite like Alter's (2008) objection to Dennett's (2005) "RoboMary" thought-experiment,

The problems with the Locked RoboMary case are symptomatic of a more general difficulty with Dennett's strategy. If the states Mary, RoboMary, or another Mary counterpart puts herself in- states that enable her to deduce what it's like to see red- involve color phenomenology, then she cheats: she does not a priori deduce the phenomenology from the physical information. In that case, her achievement fails to threaten the non-deducibility claim. If, however, the states she puts herself in do not involve color phenomenology, then it is hard to see how they would enable her to deduce the phenomenology. I see no way of modifying the RoboMary case to evade this dilemma. (Alter, 2008: 75)

Even though Hypersensitive Mary will identify red "on sight", the novelty of her experience will be significant. The way her visual system processes this polychromatic input will be significantly different from the way it has processed its former monochromatic input. If this is a substantially novel form of visual information processing for Hypersensitive Mary (and it almost certainly is), it may "train up" her neural network such that it processes polychromatic information differently when it is encountered in the future (this "future" being the very next moment in her visual processing, however that is measured). This, I believe, is that what is actually meant by the expression "knowing by sight", rather than knowing by inference. It is the difference between (1) and (2) below:

(1) The way in which Mary will process color information at t_1 , the first time she encounters a polychromatic stimulus (on the personal and sub-personal levels of information processing),

(2) How she will process color information on all subsequent encounters, t_2 - t_n .

T_1 and t_2 will both take place in the specious present, as will t_3 and t_{n-1} ; however there is a genuine distinction to be made between Mary's information-processing state at her initial encounter and such states in all subsequent encounters, regardless of whether the subsequent encounters can be given a precise chronological order. At the time of the initial encounter with a color stimulus, her visual system is in a state V , a state the system can enter only if it has never been exposed to color (a color-virginal state). Whatever the state of her visual at encounters t_2 - t_n , it will not be state V .

6.6 The “Cognitive Stance”

All sides in the debate seem to be looking at the facts Mary learns in her black and white room as physically *reduced* scientific facts, rather than physically *reducible* scientific facts- taking the *physical* stance toward Mary's knowledge rather than the more appropriate “cognitive stance”.

Given that property dualist have created an ontological divide between cognitive and phenomenal facts, all facts on the non-phenomenal side of that divide are *by hypothesis* physically reducible and therefore facts to which Mary should be privy in her studies. Whether these facts include facts about particulars or are only general facts should not matter if we are taking the cognitive stance toward Mary in her learning; since cognitive states supervene (or are instantiated in) a multitude of brain states, Mary does not need to learn to identify particular states of her brain but (what is more general) the cognitive states that may supervene upon a plethora of physical states. There will be some story, at the cognitive level,

of how Mary comes to judge that *this* is red, as well as a particular physical story (on which that cognitive story supervenes, if you will).

This is where previous attempts to explain how Mary comes to know what it's like to see red *in virtue of what she learns prior to seeing color* seem to go awry. - Mary need not put herself into a red-experiencing-state or red-experienced-neurological-state in order to know what it's like to see red (Dennett (2005) and Beaton (2005) to the contrary). The goings-on at the sub-personal level are relevant to knowing what it's like to see red only insofar as these neural events are the substrate for the *personal-level psychological state* knowing what it's like to see red. Mary- *the person*- knows what it's like to see red- Mary's *brain* only knows what it's like to see red in the same way that her brain knows the English language- metaphorically. Mary distinguishes between colors R465 and R466 not *simply* because her brain processes information differently when it encounters R465 than when it encounters R466- this difference must appear at the personal-level, psychological reaction to R465 and R466 or *Mary will not make the discrimination no matter what difference there may be in her brain activity* in the presence of R465 and R466.

This is more akin to how we actually discriminate color- whatever might be happening at the neural level when we encounter stimuli x and y, unless that neural event influences activity at the cognitive level we will not notice a difference between x and y, even if there is a consistent difference between the brain activity when we perceive red456 and when we perceive red457. Some brain event may be regularly correlated with red 456 but not 457, and vice versa. But unless this physical difference makes a difference at the cognitive level *we will not notice* a difference between perceiving red456 and red457. By this I mean that there will be *nothing that it's like* to perceive red456 *rather than* red457 if the difference

between is smaller than the minimum distinguishable difference, which is a distinction at the cognitive level, not the physical level.

Because we know that the metaphysics are irrelevant to whether Mary can identify red on sight, the real debate is now between the a priori physicalists and a posteriori physicalists. A posteriori physicalists believe that phenomenal and non-phenomenal concepts are isolated from each other, thus Mary will acquire a new phenomenal concept, Red_{ph} , when she has her first experience of redness, and will do so regardless of her ability to identify color “on sight”. A priori physicalists could dispute this for a variety of reasons, depending on their view of conceptual analysis, whether they reject or accept the notion of “pure a priori reasoning”,¹⁴² or whether they believe that our mental concepts are amalgams of cognitive, functional, and “phenomenal” features rather than “cleanly” separated along the lines of “experience” and “cognition”. Wherever their differences might lie, Hypertensive Mary gives the a posteriori and a priori physicalists a new platform on which to examine their disagreements.

¹⁴² There are “Quinean physicalists” who may object to the idea of a priority and a priori arguments, which would change the structure of the discussion substantially.

7. A Conclusion

To secure the conceivability argument, the property dualist's last line of defense was to declare physicalism to be inconceivable, or at least not as conceivable as property dualism. But the unavoidable "explanatory irrelevance" of phenomenal consciousness to explanations of human behavior allowed Hypersensitive Mary to learn all she needed to know about her own cognitive states, and have all the sensitivity she needed to identify each down to the finest discriminable detail, in order for her to identify colors on sight. This does not prove that she gains no new knowledge when she leaves her black and white room, and it does not prove that there is no "phenomenal" knowledge "above and beyond" the knowledge that allowed her to identify color on sight. But it does show that the epistemic arm of the argument for property dualism is broken, and that physicalism is *at least as conceivable* as property dualism. If the "anti-zombie" argument laid out in chapter 5 holds up to scrutiny, then together the "anti-zombies" and Hypersensitive Mary have collapsed the conceivability argument.

As for the zombie's "phenomenal" concept and "phenomenal" belief; "Optimistic Schmenomenality", the interpretation of zombie-"phenomenal" talk that gives the zombie's concepts and beliefs the same epistemic status as our own, did not fail for its optimism. Not only are Inverted Schmenomenal-Zombie-Twins conceptually incoherent- any kind of Inverted-Zombie-Twin will be incoherent. The conceptual incoherence of Inverted-Zombie-Twins is a problem for the interpretation of zombie-"phenomenal" concepts on which these

concepts refer to something that exists in the zombie world and for the interpretation of zombie-“phenomenal” concepts on which these concepts are empty.

All along we accepted that zombie “phenomenal” concepts had no phenomenal content, so that zombie “phenomenal” beliefs, whatever these beliefs were about, were not about *our kind* of phenomenal qualities. Perhaps we should reconsider that option, and allow the zombie’s propositional attitudes to have phenomenal objects, for zombie-“phenomenal” beliefs to be about real properties. But to take this route, we would need to overcome the obstacle of the zombie having neither actual nor potential epistemic access to phenomenal properties. If this obstacle is insurmountable, and both the replacement-phenomenal-referent strategy and empty concept strategy lead to creature that ought to be conceivable but are actually conceptually incoherent, we seem to have run into a wall in attempting to give a coherent account of zombie-“phenomenal” talk. We may need to consider the option that zombie-“phenomenal” talk is meaningless, but if so, we should closely examine our criteria for “meaningful language” relative to our dedication to the ideal conceivability of zombies.

Has any of this shown that zombies are inconceivable? We have uncovered some genuinely inconceivable minds- “Conceptually-Distinct Zombie-twins of Inverted-twins” and “Inverted-Zombie-twins”. But stories about “The Zombie” and zombie-scenarios will always be intelligible; we must simply bear in mind that intelligibility alone is not significant in the debate over the metaphysics of mind. Intuition is the great instigator of philosophical thought, but if philosophy is to move beyond bare intuition into serious investigation, the heavy lifting must be handed over to our faculty of Reason.

Bibliography

- Armstrong, D. (1963). *A Materialist Theory of Mind*. London: Routledge & Kegan Paul.
- Alter, T. (2007). "Phenomenal Knowledge without Phenomenal Experience." In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Alter, T. & S. Walter (eds.) (2007). *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Balog, K. (1999). "Conceivability, possibility, and the Mind-Body Problem." *Philosophical Review*. 108: 497-528.
- . (2008). "Illuminati, Zombies, and Metaphysical Gridlock." (MS).
- . (2009). "Phenomenal Concepts." In B. McLaughlin, A. Beckermann & S. Walter (eds.) *Oxford Handbook in the Philosophy of Mind*. Oxford: Oxford University Press.
- . (2011). "Acquaintance and the Mind-Body Problem." In C. Hill & S. Gozzano (eds.) *The Mental and the Physical*. Cambridge: Cambridge University Press.
- Bealer, G. (1997). "Self-Consciousness." *Philosophical Review* 106: 69-117.
- Beaton, M. (2005). "What RoboDennett Still Doesn't Know." *Journal of Consciousness Studies* Vol. 12. No. 12: 3-25.
- Beckermann, A., H. Flohr & J. Kim (eds.) (1992). *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin & New York: De Gruyter.
- Black, M. (ed.) (1965). *Philosophy in America*. London: Routledge and Kegan Paul.
- Block, N. (1978). "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9: 261-325.
- . (ed.) (1980). *Readings in Philosophy of Psychology*, vol. I. Cambridge: Harvard University Press.
- . (1990). "Inverted Earth." In J. Tomberlin (ed.) *Philosophical Perspectives* 4. Northridge: Ridgeview Publishing Company.
- . (1995). "On a Confusion about the Function of Consciousness." *Behavioral and Brain Sciences*, 18 (2): 227-287.
- . (1997). "On a Confusion about a Function of Consciousness." In N. Block, O. Flanagan, & G. Guzeldere (eds.), *The Nature of Consciousness*. Cambridge, MA:

- MIT Press, 1997.
- . (2002). “The Harder Problem of Consciousness.” *The Journal of Philosophy* XCIX, No. 8, August 2002, 1-35.
- . (2003). “The Harder Problem of Consciousness” (longer version). *Disputatio* 15: 5-49.
- . (2007). “Max Black’s Objection to Mind-Body Identity,” In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Boghossian, P. (1997). “What the Externalist Can Know A Priori.” *Proceedings of the Aristotelian Society* 97, 161-175.
- . (1998). “Replies to Commentators.” In E. Villanueva (ed.) *Philosophical Issues 9: Concepts*: 253-260.
- Botterell, A. (2001). “Conceiving What is Not there.” *Journal of Consciousness Studies* (8): 21-42.
- Brook, A., ed. (2005). *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge: Cambridge University Press.
- Butler, R. J., ed. (1963). *Analytical Philosophy, Second Series*. Oxford: Blackwell.
- Byrne, A. (2001). “Intentionalism Defended.” *Philosophical Review* 110: 199-240.
- Campbell, K. (1970). *Body and Mind*. London: Macmillan.
- . (1984). *Body and Mind* (2nd ed.). Notre Dame, IN: University of Notre Dame Press.
- Carruthers, P. (2004). “Phenomenal Concepts and Higher-Order Experiences.” *Philosophy and Phenomenological Research* 68.2: 316-336.
- Carruthers, P. & Veillet, B. (2007). “The Phenomenal Concept Strategy.” *Journal of Consciousness Studies* 14.9-10: 212-236.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- . (2002). “Does Conceivability Entail Possibility?” In T. S. Gendler & J. Hawthorne (eds.) *Conceivability and Possibility*, Oxford, Clarendon Press.
- . (2003a). “The Content and Epistemology of Phenomenal Belief”, in A. Jokic & Q. Smith. (eds.) *Consciousness - New Philosophical Perspectives*. Oxford: Oxford University Press.

- . (ed.) (2003b). *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- . (2006). "Two-Dimensional Semantics," In E. Lepore & B. Smith (eds.) *Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press.
- . (2007). "Phenomenal Concepts and the Explanatory Gap," In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- . (2010). *The Character of Consciousness*. Oxford: Oxford University Press.
- Churchland, P. (1989). "Knowing Qualia. A Reply to Jackson." In *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.
- Conee, E. (1994). "Phenomenal Knowledge." *Australasian Journal of Philosophy* 72.2: 136-150.
- Cottrell, A. (1999). "Sniffing the Camembert: On the Conceivability of Zombies." *Journal of Consciousness Studies*. Vol. 6. No. 1. 4-12.
- Davidson, D. (1987). "Knowing One's Own Mind." *Proceedings of the American Philosophical Association* 61: 441-458.
- Dennett, D.C. (1988). "Quining Qualia." In A. Marcel & E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford: Oxford University Press, 43-77.
- . (1990). "Quining Qualia." In W.G. Lycan (ed.) *Mind and Cognition*. Oxford: Blackwell Press, 519-548.
- . (1991). *Consciousness Explained*. Boston: Little, Brown, & Co.
- . (1995). "The Unimagined Preposterousness of Zombies." *Journal of Consciousness Studies* 2.4: 322-326.
- . (1999). "The Zombic Hunch: Extinction of an Intuition?" *Royal Institute of Philosophy Millennial Lecture*.
- . (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press/Bradford Books.
- . (2006). "Higher Order Truths about Chmess." *Topoi: An International Review of Philosophy* 25.1-2: 39-41.
- . (2007). "What RoboMary Knows." In T. Alter & S. Walter (eds.) *Phenomenal*

- Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Descartes, R. (1644/1911). *The Principles of Philosophy*, E. Haldane & G. Ross (trans.). Cambridge: Cambridge University Press.
- Dretske, F. (1996). "Phenomenal Externalism, or if Meanings Ain't in the Head, Where are Qualia?" *Philosophical Issues* 7: 143-158.
- Farrell, B. A. (1950). "Experience." *Mind* 59: 170-98
- Feigl, H. (1957). "Russell and Schlick: A Remarkable Agreement of a Monistic Solution of the Mind-Body Problem." *Erkenntnis: An International Journal of Analytic Philosophy* 9: 11-34.
- Fodor, J. (1965). "Explanation in Psychology." In M. Black (ed.) *Philosophy in America*. London: Routledge and Kegan Paul.
- Frege, G. (1892/1980). "On Sense and Reference." In P. Geach & M. Black, (eds. and trans.) *Translations from the Philosophical Writings of Gottlob Frege* (3rd Edition). Oxford: Blackwell.
- Gendler, T. & J. Hawthorne (eds.) (2005). *Perceptual Experience*. Oxford: Oxford University Press.
- Goldman, A. (1993). "The Psychology of Folk Psychology." *Behavioral and Brain Sciences* Vol. 16: 15-28.
- Griffin, D. R. (1998). *Unsnarling the World-Knot*. Berkeley, CA: University of California Press.
- Hahn, M. & B. Ramberg (eds.) (2003). *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT Press.
- Hameroff, S.R., A.W. Kaszniak, & A.C. Scott (eds.) (1998). *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*. Cambridge, MA: MIT Press/Bradford Books.
- Hardin, C. (1992). "Physiology, Phenomenology, and Spinoza's True Colors." In A. Beckermann, H. Flohr & J. Kim (eds.) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin & New York: De Gruyter.
- Harman, G. (1973). *Thought*. Princeton: Princeton University Press.
- . (1990). "The Intrinsic Quality of Experience." In J. Tomberlin (ed.) *Philosophical Perspectives* 4. Atascadero, CA: Ridgeview Publishing.

- Harman, G. & D. Davidson (eds.) (1972). *Semantics of Natural Language*. Dordrecht-Holland/Boston-USA: Reidel.
- Harnad, S. (1995). "Why and How We Are Not Zombies." *Journal of Consciousness Studies* 1: 164-167.
- Hasker, W. (1999). *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hawthorne, J. (2001a). "Intrinsic Properties and Natural Relations," *Philosophy and Phenomenological Research* 63.2: 399-403.
- . (2001b). "Causal Structuralism" *Noûs-Supplement: Philosophical Perspectives* 15: 361-378.
- Hill, C. & McLaughlin, B. (1999). "There are Fewer Things in Reality than are Dreamt of in Chalmers's Philosophy." *Philosophy and Phenomenological Research* 59: 445-454.
- Hill, C. & S. Gozzano (eds.) (2011). *The Mental and the Physical*. Cambridge: Cambridge University Press.
- Ismael, J. (1999). "Science and the Phenomenal." *Philosophy of Science* 66.3: 351-369.
- Jackson, F. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-136.
- . (1998). "Postscript on Qualia." *Mind, Method and Conditionals*. London: Routledge.
- . (2003). "Narrow Content and Representation--Or Twin Earth Revisited." *Proceedings and Addresses of the American Philosophical Association*. 77. 2: 55-70.
- Jokic, A. & Smith, Q. (eds.) (2003). *Consciousness - New Philosophical Perspectives*. Oxford: Oxford University Press.
- Kant, I. (1787/1929). *Critique of Pure Reason*, N. Kemp-Smith (trans.). New York: MacMillan.
- Kind, A. (2001). "Qualia Realism." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 104. 2: 143-162.
- Kirk, R. (1974a). "Sentience and Behaviour." *Mind* 83: 43-60.
- . (1974b). "Zombies vs. Materialists." *Proceedings of the Aristotelian Society*, Supplementary Vol. 48: 135-152.
- . (2005). *Zombies and Consciousness*. Oxford: Clarendon Press.
- . (2008). "Zombies and Consciousness." In Y. Nagasawa (ed.) *Philosophical Books*

49. 2: 170-171.
- Kripke, S. (1972). "Naming and Necessity." In G. Harman and D. Davidson (eds.) *Semantics of Natural Language*. Dordrecht-Holland/Boston-USA: Reidel: 253-355.
- . (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lamme, V. A. (2006). "Towards a True Neural Stance on Consciousness." *Trends in Cognitive Science*. Nov. 10(11): 494-501.
- Leibniz, G.W. (1720/1925). *The Monadology*, D. Garter & R. Aries (trans.) Indianapolis: Hackett.
- Lepore, E. & B. Smith (eds.) (2006) *Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press.
- Levin, J. (2007). "What Is a Phenomenal Concept?" In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Levine, J. (1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly*. 64. 354-361.
- . (2001). *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.
- . (2007). "Phenomenal Concepts and the Materialist Constraint." In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Lewis, C.I. (1929). *Mind and the World Order: Outline of a Theory of Knowledge*. New York: Charles Scribners.
- Lewis, D. (1966). "An Argument for the Identity Theory." *Journal of Philosophy* 63: 12-25.
- . (1980). "Mad Pain and Martian Pain." In N. Block (ed.) *Readings in Philosophy of Psychology*, vol. I. Cambridge: Harvard University Press, 216-232.
- . (1983). "Extrinsic Properties." *Philosophical Studies* 44: 197-200.
- . (1988). "What Experience Teaches." *Proceedings of the Russellian Society*, University of Sydney, 13: 29-57.
- Loar, B. (1990). "Phenomenal States." *Philosophical Perspectives 4: Action Theory*: 81-108.
- . (1997). "Phenomenal States." (second version), In N. Block, O. Flanagan, & G.

- Güzeldere (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- . (2003). "Phenomenal Intentionality as the Basis of Mental Content," in M. Hahn & B. Ramberg, (eds.) *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT Press.
- Locke, J. (1688/1959). *An Essay on Human Understanding*. New York: Dover.
- Lowe, E.J. (1995). *Locke on Human Understanding*. London: Routledge.
- Lycan, W.G. (ed.) (1990). *Mind and Cognition*. Oxford: Blackwell Press.
- Lycan, W.G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- . (2008). "Phenomenal Intentionalities." *American Philosophical Quarterly* 45 (3): 233-252.
- Maloney, J.C. (1985). "About Being a Bat." *Australasian Journal of Philosophy* 63.1: 26.
- Marcel, A. & E. Bisiach (eds.) (1988). *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- Marcus, E. (2004). "Why Zombies are Inconceivable." *Australasian Journal of Philosophy* 82.3: 477-490.
- Matey, J. (2012). "Representing the Impossible." Forthcoming in *Philosophical Psychology*.
- Maund, B. (2008). "A Defense of Qualia in the Strong Sense." In E. Wright (ed.) *The Case for Qualia*. Cambridge, MA: MIT Press/Bradford Books.
- McGeer, V. (2003). "The Trouble with Mary." *Pacific Philosophical Quarterly* 84.4: 384-393.
- McGinn, C. (1989). "Can We Solve the Mind-Body Problem?" *Mind* 98: 346-366.
- McLaughlin, B., A. Beckermann & S. Walter (eds.) (2009). *Oxford Handbook in the Philosophy of Mind*. Oxford: Oxford University Press.
- Metzinger, T. (ed.) (1995). *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- Moore, G.E. (1903). "The Refutation of Idealism." *Mind* 12: 433-453.
- Murdoch, I. (1970/2002). *The Sovereignty of the Good*. London/New York: Routledge.
- Nagel, T. (1974). "What is it like to be a Bat?" *Philosophical Review* 83: 435-456.
- Nemirow, L. (1990). "Physicalism and the Cognitive Role of Acquaintance." In W.G. Lycan

- (ed.) *Mind and Cognition: A Reader*. Oxford: Blackwell: 490-499.
- . (2007). “So *This* is What it’s Like: A Defense of the Ability Hypothesis.” In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Nida-Rumelin, M. (1995). “What Mary Couldn’t Know: Belief about Phenomenal States.” In Metzinger, T. (ed.) *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- O’Hear, A., (ed.) (2003). *Minds and Persons*. Cambridge: Cambridge University Press.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- . (2007). “Phenomenal and Perceptual Concepts.” In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Peacocke, C. (1983). *Sense and Content*. Oxford: Oxford University Press.
- Penrose, L.S. & Penrose, R. (1958). “Impossible objects: A special type of visual illusion.” *British Journal of Psychology* 49: 31-33
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.
- Place, U.T. (1956). “Is Consciousness a Brain Process?” *British Journal of Psychology*: 44-50.
- Prinz, J. (2005). “A Neurofunctional Theory of Consciousness.” In A. Brook (ed.) *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge: Cambridge University Press, 381-396.
- . (2006). “Beyond Appearances: The Content of Sensation and Perception.” In T. Szabó (ed.) *Perceptual Experience*. Oxford: Clarendon Press, 434-460.
- . (2007). “Mental Pointing: Phenomenal Knowledge without Concepts.” *Journal of Consciousness Studies* 14.9-10: 184-211.
- Putnam, H. (1963). “Brains and Behavior.” In R. J. Butler (ed.) *Analytical Philosophy, Second Series*. Oxford: Blackwell, 211-235.
- Raffman, D. (2005). “Even Zombies Can Be Surprised: A Reply to Graham and Horgan.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 122.2: 189-202.
- Rey, G. (1998). “A Naturalistic A Priori.” *Philosophical Studies: An International Journal for the Philosophy in the Analytic Tradition* 92.1-2: 25-43.

- Robinson, H. (1994). *Perception*. London: Routledge.
- Robinson, W.S. (1988). *Brains and People: An Essay on Mentality and Its Causal Conditions*. Philadelphia: Temple University Press.
- . (2004). *Understanding Phenomenal Consciousness*. Cambridge: Cambridge University Press.
- . (2012). "Epiphenomenalism." In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, <<http://plato.stanford.edu/entries/epiphenomenalism/>>.
- Rorty, A. (ed.) (1976). *The Identities of Persons*. Berkeley: University of California Press.
- Rosenberg, G. (2004). *A Place for Consciousness*. Oxford: Oxford University Press.
- Russell, B. (1912/1997). *The Problems of Philosophy*. Oxford: Oxford University Press.
- Ryle, G. (1949/2000). *The Concept of Mind*. London: Penguin Books.
- Seager, W. (1991). *Metaphysics of Consciousness*. London: Routledge.
- Searle, J. (1990). "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3: 417-457.
- . (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Sellars, W. (1965). "The Identity Approach to the Mind-Body Problem." *Review of Metaphysics* 18: 430-451.
- Shoemaker, S. (1975). "Functionalism and Qualia." *Philosophical Studies* 27: 291-315.
- . (1976). "Embodiment and Behavior." In A. Rorty (ed.) *The Identities of Persons*. Berkeley: University of California Press.
- . (1994). "Phenomenal Character." *Noûs* 28.1: 21-38.
- Siewert, C. (1998). *The Significance of Consciousness*. Princeton: Princeton University Press.
- Smart, J.J.C. (1971). "Reports of Immediate Experiences." *Synthese* 22.3-4: 346-359.
- Smythies, J. (2008). "The Ontological Status of Qualia and Sensations: How They Fit into the Brain." In E. Wright (ed.) *The Case for Qualia*. Cambridge, MA: MIT Press/Bradford Books.
- Stoljar, D. (2005). "Physicalism and Phenomenal Concepts." *Mind and Language* 20.2: 296-302.

- Strawson, G. (1994). *Mental Reality*. Cambridge, MA: MIT Press/Bradford Books.
- Stubenberg, L. (1998). *Consciousness and Qualia*. Amsterdam: John Benjamins.
- Szabó, T. (ed.) (2006). *Perceptual Experience*. Oxford: Clarendon Press.
- Taliaferro, C. (1996). *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.
- Thomas, N. J. T. (1998). "Zombie Killer." In S.R. Hameroff, A.W. Kaszniak, & A.C. Scott (eds.) *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*. Cambridge, MA: MIT Press/Bradford Books.
- Tye, M. (2000). *Color, Content, and Consciousness*. Cambridge, MA: MIT Press/Bradford Books.
- . (2003a). "A Theory of Phenomenal Concepts." In A. O'Hear (ed.) *Minds and Persons*. Cambridge: Cambridge University Press.
- . (2003b). "Visual Qualia and Visual Content Revisited." In D. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- . (2005). "Nonconceptual Content, Richness, and Fineness of Grain." In T. Gendler & J. Hawthorne (eds.) *Perceptual Experience*. Oxford: Oxford University Press.
- . (2009). "A New Look at the Speckled Hen." *Analysis* 69.2: 258-263.
- Vierkant, T. (2002). "Zombie-Mary and the Blue Banana: On the Compatibility of the 'Knowledge Argument' with the Argument from Modality." *Psyche: An Interdisciplinary Journal of Research on Consciousness* 8.19.
- Villanueva, E. (ed.) (1996). *Philosophical Issues, 7: Perception*. Atascadero, CA: Ridgeview Publishing.
- White, S. L. (2007). "Property Dualism, Phenomenal Concepts, and the Semantic Premise." In T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays in Consciousness and Physicalism*. Oxford: Oxford University Press.
- Wilson, R. A., Keil, F. C. (eds.) (2001). *MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*. G.E.M. Anscombe & R. Rhees (eds.) G.E.M. Anscombe (trans.) Oxford: Blackwell.
- Wright, E. (ed.) (2008). *The Case for Qualia*. Cambridge, MA: MIT Press/A Bradford Book.

Yablo, S. (1993). "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53.1: 1-42.

_____. (1999). "Concepts and Consciousness." *Philosophy and Phenomenological Research* 59.2: 455-463.

_____. (2003). "Causal Relevance." *Noûs-Supplement: Philosophical Issues* 13: 316-328.