

**UNIVERSITA' VITA-SALUTE SAN RAFFAELE**

**CORSO DI DOTTORATO DI RICERCA**

**IN FILOSOFIA**

**Curriculum in Philosophical Issues in an International Perspective**

**CONSCIOUSNESS, PRESENTATION,  
AND (SELF)REPRESENTATION**

Tutore: Prof. Elisabetta Sacchi



Co-Tutore: Dr. Sam Coleman

Tesi di DOTTORATO DI RICERCA di Davide Zottoli

matr. 015516

Ciclo di Dottorato XXXV

SSD M-FIL/05

Anno Accademico 2021/2022



## CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il sottoscritto Davide Zottoli

Matricola / *registration number* 015516

nato a/ *born at* Torino

il/on 28/12/1994

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

“Consciousness, Presentation, and (Self)Representation”

AUTORIZZA la Consultazione della tesi / *AUTHORIZES the public release of the thesis*

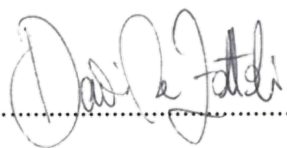
NON AUTORIZZA la Consultazione della tesi per ..... mesi /*DOES NOT AUTHORIZE the public release of the thesis for ..... months*

a partire dalla data di conseguimento del titolo e precisamente / *from the PhD thesis date, specifically*

Dal / *from* ...../...../..... Al / *to* ...../...../..... Poiché / *because:*

- l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*
- ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*
- la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data / *Date* ..21/01/2023..... Firma / *Signature* .......

## **DECLARATION**

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.
- has been written according to the editing guidelines approved by the University.

All the results presented here were obtained by myself.

All sources of information are acknowledged by means of reference.

Parts of this thesis (§4) have been published before. Details of the original publication follow.

Zottoli D, (2022) Intentionality and Inner Awareness. *Phenomenology and Mind* 22: 68-81.

## **ACKNOWLEDGEMENTS**

Completing a Ph.D. Program turned out to be a far more difficult task than I could have ever imagined – one that I could have never accomplished without the invaluable help I received, and thus that could not be considered quite finished without my thanks to those who have helped me along the way.

First, I would like to thank my tutor, Elisabetta Sacchi, for her guidance and understanding in these strange last few years, and my co-tutor, Sam Coleman, for his kindness and hospitality, and the countless insightful conversations about the topics of this dissertation and much more.

Second, I would like to thank David Rosenthal, whose comments and suggestions greatly improved my understanding of the many longstanding debates I encountered, and Alberto Voltolini, who first sparked my interest in the philosophical complexities of the mind and without whom this Ph.D. journey may have never started.

Finally, I would like to thank my family, for their support every step of the way and for always allowing me, in the first place, to freely try becoming who I chose to be; and I would like to thank my friends, for saving my sanity during the pandemic and for the many ways in which they help me appreciate life.

## **ABSTRACT**

The purpose of this dissertation is to provide an analysis of the possible approaches to the hard problem of consciousness within the framework of higher-order intentionalism, and to argue for the superiority of extrinsic higher-order theories – according to which inner awareness is a cognitive mechanism distinct from phenomenal character – over intrinsic higher-order theories – according to which it is a mental state’s acquiring a phenomenal character that is responsible for the constitution of inner awareness. In Part I, it will be argued that intrinsic higher-order theories involve controversial metaphysical assumptions as well as the commitment to the contentious claim that consciousness should be conceived as being primarily a property of mental states. In Part II, by analysing the fundamental dimensions of variation among specific higher-order theories, it will be argued that while intrinsic higher-order theories may appear appealing in virtue of their promise to provide a solution to the hard problem, it is doubtful that they can deliver on such a promise, and that the fundamental principle behind extrinsic higher-order theories (traditionally associated with the rejection of the hard problem) can be developed into promising higher-order theories able to acknowledge the reality of hard problem.

L’obiettivo di questa dissertazione è offrire un’analisi dei possibili approcci al problema della coscienza offerti dall’intenzionalismo di second’ordine, e di sostenere la superiorità delle versioni estrinseche di tale teoria – secondo cui la coscienza consiste in un meccanismo cognitivo distinto dal carattere fenomenico degli stati mentali coscienti – rispetto alle versioni intrinseche – secondo cui la consapevolezza dei propri stati mentali è costituita da proprietà che sono parte integrante del carattere fenomenico degli stati coscienti. Questa tesi verrà difesa, nella prima parte della tesi, sostenendo che le teorie intrinseche presuppongono tesi controverse in ambito metafisico e una discutibile caratterizzazione della coscienza intesa principalmente come una proprietà degli stati mentali e, nella seconda parte della tesi, analizzando le differenze tra specifiche formulazioni di tali teorie e sostenendo che da un lato le teorie intrinseche non siano in grado di mantenere la loro promessa di risolvere il problema della coscienza, mentre dall’altro lato i principi delle teorie estrinseche (tradizionalmente associate al diniego di tale problema) possono essere utilizzati per elaborare soluzioni al problema della coscienza significativamente più promettenti.

## **Table of contents**

<b>Introduction</b>	1
<b>Part I. Conscious states and Conscious Subjects</b>	4
Introduction	4
<b>1. The Phenomenal Character View</b>	8
1.1. Consciousness as Phenomenal Character	10
1.1.1. Varieties of intrinsic qualities	10
1.1.2. Intrinsic qualities and the hard problem	14
1.2. Approaches to the Hard Problem	18
1.2.1. Eliminativism	18
1.2.2. Illusionism	23
1.2.3. Realism	26
<b>2. The Extrinsic View</b>	28
2.1. Varieties of Extrinsic Views	28
2.2. Varieties of Phenomenal Properties	39
<b>3. Metaphysical Implications</b>	51
3.1. Consciousness as Categorical or Dispositional	51
3.1.1. Categoricalism and dispositionalism	51
3.1.2. Dispositionalism and the categorical nature of phenomenal character	54
3.2. Consciousness as Categorical and Dispositional	62
3.2.1. Identity views	63
3.2.2. Dualism	67
3.2.3. Mixed dispositionalism	72
<b>Conclusions to Part I</b>	78
<b>Part II. Varieties of Higher-Order Theories</b>	81
Introduction	81
<b>4. Representationalist Higher-Order Theories</b>	<b>82</b>
4.1. Extrinsic Theories	83

4.1.1.	HOT vs. HOP	83
4.1.2.	The objection from distinctness	90
4.1.3.	The generality problem	98
4.2.	Intrinsic Theories	102
4.3.	The Implications of Representationalism	107
4.3.1.	The objection from intimacy and the constituting representation view	107
4.3.2.	Representationalism and illusionism	116
<b>5.</b>	<b>Higher-Order Intentionalism and Realism about the Hard Problem</b>	<b>121</b>
5.1.	Quotational Higher-Order Thoughts	123
5.2.	The Subject View and a HOP-like Alternative	136
5.2.1.	Higher-order global states	137
5.2.2.	The attention schema: from illusionism to realism	146
	<b>Conclusions</b>	<b>156</b>
	<b>References</b>	<b>161</b>



## Introduction

The purpose of this dissertation is to provide an analysis of the possible approaches to the hard problem of consciousness<sup>1</sup> within the framework of higher-order intentionalism,<sup>2</sup> and to argue for the superiority of extrinsic higher-order theories – according to which consciousness is distinct from, and at least partly responsible for the constitution of phenomenal character<sup>3</sup> – over intrinsic higher-order theories – according to which it is a mental state’s acquiring a phenomenal character that is responsible for the constitution of consciousness.<sup>4</sup>

The first part of the dissertation will be focused on the contrast between the conception of consciousness involved in the formulation of extrinsic theories and the conception of consciousness presupposed by intrinsic theories, and it will be argued that the former offers to the supporter of higher-order intentionalism significant advantages over the latter – in that it allows to adopt a wider variety of explanatory strategies to tackle the hard problem and to avoid controversial metaphysical commitments. The second part of the dissertation will be devoted to the analysis of the fundamental dimensions of variation among specific higher-order theories, and it will be argued that the conception of consciousness adopted within the framework of extrinsic theories is considerably more fruitful than the one proposed within the framework of intrinsic theories: although the latter is naturally seen as promising to provide a solution to the hard problem, it is doubtful that it can succeed; and while the former does not require taking the hard problem at face value (and has been traditionally associated with its rejection), it can offer the most promising higher-order strategies to face it.

Part I consists of three chapters. The first chapter focuses on the conception of consciousness presupposed by intrinsic higher-order theories: the ‘phenomenal character

---

<sup>1</sup> That is, the problem of explaining why the physical phenomena responsible for the performance of cognitive and behavioural functions are accompanied by phenomenally conscious experience, or mental states endowed with phenomenal properties – and, relatedly, why each of those physical phenomena is accompanied by a specific type of experience rather than another (Chalmers 1996).

<sup>2</sup> According to higher-order intentionalism, the existence of consciousness should be explained in terms of the subject’s inner awareness of her mental states, which in turn should be characterized in terms of (higher-order) intentionality (i.e., the property in virtue of which mental states can exhibit *directedness*, or *aboutness* towards some (intentional) object, property, or state of affairs).

<sup>3</sup> The notion of phenomenal character will be defined as referring to the sum of the phenomenal properties of the mental states becoming conscious that determine the phenomenal contents of conscious experience.

<sup>4</sup> The definition of the notions of intrinsic and extrinsic that will be adopted here relies on the contrast between something’s having some properties “in virtue of the way that thing itself, and nothing else, is” (Lewis 1983, 197) and other properties in virtue of the way it is related to other entities – in a way analogous to Kriegel (2009, 145).

view’, according to which phenomenal consciousness is made of essentially conscious properties of mental states that make their subject conscious. After a brief presentation of this conception of consciousness, it will be argued that the phenomenal character view directly implies that consciousness is constituted by intrinsic properties of conscious states (hence the name, *intrinsic* higher-order theories) and that for this reason it naturally leads to the formulation of hard problem. The remainder of the chapter will be focused on the presentation of the ontological and explanatory differences between the phenomenal character view and (extrinsic) conceptions of consciousness that reject the reality of the hard problem.

The second chapter focuses on the conception of consciousness presupposed by extrinsic higher-order theories: the ‘extrinsic view’, according to which phenomenal consciousness is made of non-essentially conscious properties of mental states that are made conscious by cognitive mechanisms that are not parts of the phenomenal contents of conscious experience. After presenting the fundamental types of explanatory strategies available to the higher-order theorist within this framework, it will be argued that despite their traditional association with the rejection of the hard problem, extrinsic views are in principle compatible with a realist attitude towards the hard problem – either by conceiving consciousness as being primarily a property of mental states (while taking inner awareness to be necessary but not sufficient for the constitution of consciousness), or by conceiving consciousness as being primarily a property of subjects (a choice that allows taking inner awareness to be sufficient for the constitution of consciousness). The remainder of the chapter will be devoted to the clarification of this latter, rather unorthodox view.

The third chapter offers an analysis of the metaphysical commitments involved in the alternative conceptions of consciousness presented in the previous chapters. It will be argued that the extrinsic view may be preferred over the phenomenal character view because of its metaphysical neutrality – contrasting with the phenomenal character view’s commitment to a categorialist conception of consciousness: on intrinsic higher-order theories, any dispositional feature of consciousness (if there are any) must be grounded on a categorial basis, while extrinsic theories can allow to conceive of consciousness as being a dispositional property, though they are also compatible with categorialism.

Part II consists of two chapters. The fourth chapter focuses on representationalist versions of higher-order intentionalism, according to which inner awareness is constituted by the representations of one's own first-order states. After presenting the main types of extrinsic higher-order representationalism and assessing the main objections devised against them, it will be argued that although intrinsic higher-order representationalism may appear as being better equipped to avoid those same objections, it ultimately shares its fate with extrinsic higher-order representationalism – as both inevitably lead to deny that there is a hard problem of consciousness.

Finally, the fifth chapter provides an overview of the strategies available to the higher-order theorist for taking the hard problem at face value, thereby concluding the argument for the superiority of extrinsic views: the explanatory strategy adopted by extrinsic higher-order theories is considerably more fruitful than the one proposed by supporters of intrinsic higher-order theories because, once representationalism is dropped, the latter leads to abandon higher-order intentionalism whereas the former offers a variety of viable explanations.

## **Part I. Conscious states and Conscious Subjects**

### **Introduction**

After the decline in popularity of behaviourism and the rise of cognitive sciences – bridging the domains of folk-psychology and neurosciences by conceiving the central nervous system as a implementing a complex functional architecture consisting of (possibly unconscious) mental states endowed with intentional content, or *aboutness* – the quest for a naturalistic understanding of consciousness slowly returned to the spotlight in the English-speaking world, within both philosophical and scientific circles (Armstrong 1980; Rosenthal 1986; Lycan 1987; Baars 1988; Crick & Koch 1990; Dennett 1991; Dretske 1995; Tye 1995).

Understanding consciousness within a naturalistic framework means being able to explain – consistently with the causal closure of the physical world – the reason why some physical phenomena responsible for the performance of cognitive and behavioural functions are accompanied by an experienced inner life, rather than feeling like nothing at all, and why the experiencing associated with each of those phenomena feels exactly like *this* rather than like *that*. On the one hand, one must be able to explain why, for example, there is something it is like to see the blue sky, to hear the waves crashing on the shore, to taste the sweetness of an apple, to smell some freshly cut grass and to cut it: couldn't every bit of neural activity underlying each of those experiences take place “in the dark”? On the other hand, one must also be able to explain why, for example, consciously perceiving the sky presents one with a certain specific bluish quality rather than another, why the waves crashing on the shore do not sound like the cry of a hungry seagull, and why freshly cut grass does not smell like freshly brewed coffee nor does it have the sweetish taste of an apple, and so on.

The now-mainstream (though not uncontroversial) idea that the mind should be conceived as an information-processing system naturally led most of the philosophers listed above (and many others) to suppose that the identity of a conscious experience is determined by, or at least supervenient on its intentional properties, i.e., what the experience is *about* and by the way in which it represents it. Thus, for example, the bluish quality perceived when consciously seeing the sky is precisely the quality it is because the subject is in a conscious visual state that represents the blueness of sky (with specific

degrees of hue, saturation and value).<sup>5</sup> But then, once again, it is natural to ask: why couldn't that visual representation of the sky occur unconsciously? Many have tried to answer this question by identifying a specific functional role that could be the distinguishing factor between conscious and unconscious representations, thereby determining the existence conditions of conscious experience. For example, according to Baars (1988), the information represented by a certain mental state becomes a content of conscious experience when it is made cognitively available for widespread access and use in virtue of being broadcasted to a global workspace – a central processor working as a medium of communication for more specialized information-processing systems; according to Crick & Koch (1990), conscious mental states are those that go through a 'binding' process (allegedly, in virtue of their neural realizer's synchronization at a 35-75 hertz oscillation frequency), by means of which pieces of information represented by independent sub-systems but concerning one and the same entity are brought together to be used by later processing; and according to Dretske (1995) and Tye (1995), for a mental state to be conscious is for it to be poised to be used in cognitive processing that produces beliefs and desires and allows flexible control of behaviour. Yet, for any such proposal, it is still possible to ask: why should the performance of that function give rise to conscious experience? It seems that a mental state may be poised for use in cognitive processing, or have its intentional content properly binded, or be globally accessible, and yet, at the same time, remain unconscious (e.g., when being jealous of someone without being aware of it, or when one's moods and behaviours are constantly affected by a chronic pain even though the feeling of pain goes in and out of consciousness). A pessimist may conclude that looking for a cognitive function such that its explanation will automatically also explain the existence of conscious experience is simply a hopeless quest (Levine 1983; Block 1995; Chalmers 1996). There is, however, a more hopeful alternative: one may try to explain what distinguishes conscious and unconscious states by applying the intentionalist treatment, rather than (only) to *what* conscious states present to their subject, also to *how* they present it. The intuitive ground behind this strategy is the likely observation that there is nothing it is like to be in a mental state

---

<sup>5</sup> It is not clear whether all experienced qualities are in fact associated with a represented property (e.g., Kind 2003), nor whether it is possible to reduce qualities to intentional contents, as phenomenal notions may be needed for individuating them and for determining the relevant mode of representation (e.g., Chalmers 2004). However, these questions will be overlooked here, since this dissertation is focused on the question concerning the existence conditions of conscious experience, rather than its identity conditions.

whose experienced qualities are not subjectively given, and the plausible suggestion that the subjective givenness of conscious experiences may be constituted by the subject's awareness of *being in* those mental states. That is, there being something it is like for a subject to be in a mental state (say, a visual perception of something red) is not only "a matter of some state (my experience) having some feature (being reddish)", because it also depends on the fact that "its being reddish is 'for me'" (Levine 2001, 6-7); and in turn this essentially subjective aspect of conscious experience may be construed in terms of the subject's instantiation of higher-order intentional properties directed at the mental states subjectively experienced. This strategy is synthesized by the so-called 'transitivity principle', according to which "conscious states are simply mental states we are conscious of being in" (Rosenthal 1986, 26), and naturally leads to the formulation of the higher-order intentionalist account of consciousness – defined by the conjunction of two theses: that the existence of consciousness depends on the subject's inner awareness of her mental states, and that inner awareness is constituted by higher-order intentionality.<sup>6</sup>

Higher-order intentionalism has been traditionally distinguished from its competitors by means of the contrast between conceptions of conscious states as "states that we make conscious by being conscious of them" and conceptions of conscious states as "states we are conscious *with*", i.e., "states that make us conscious" (Dretske 1995, 100-1). The basic idea behind this distinction is that the higher-order theorist suggests that consciousness should not be characterized solely in terms of 'outer' awareness, i.e., concerning our awareness of the world around us (including ourselves, as part of that world), because it essentially involves the inner awareness of one's own mental life. However, the distinction between these two conceptions of conscious states can also be applied within the domain of higher-order intentionalism, in order to distinguish two radically different senses in which the constitution of consciousness can be said to depend on the presence of inner awareness. On the one hand, inner awareness may be conceived as a property that is distinct from the experienced properties of the mental state made conscious (and

---

<sup>6</sup> According to some philosophers (e.g., Block 1995), this kind of cognitive access can only account for a type of consciousness that is distinct from the topic of this dissertation, i.e., *phenomenal* consciousness. Yet, this thesis is certainly more controversial than the criticisms of the functional criteria considered above, insofar as there is an answer to the question of why the performance of the cognitive functions responsible for inner awareness is accompanied by conscious experience: because endowing a subject with inner awareness makes him aware of being in certain mental states (in a way that does not apply to unconscious states). Those who are not persuaded that this is a viable answer may read this dissertation as an attempt to find specific formulations of it that may be more promising than traditional ones.

at least partly responsible for our experiencing them), such that a certain mental state is made conscious by the presence of the subject's inner awareness (e.g., Armstrong 1980; Lycan 1987; Rosenthal 2005, Coleman 2015). On the other hand, inner awareness may be conceived as a property that is constituted by the experienced properties of the mental state that becomes conscious – such that it is not the presence of inner awareness that explains a mental state's becoming conscious but, rather, it is a state's becoming conscious that establishes the relation of inner awareness with the subject (e.g., Carruthers 2000, Caston 2002; Williford 2006; Kriegel 2009). While the former view of inner awareness leads to the formulation of 'extrinsic' higher-order theories, according to which conscious states are strictly speaking mental states that are made conscious by the subject's having inner awareness, the latter view leads to the formulation of 'intrinsic' higher-order theories that blend the two horns of Dretske's distinction together, in that conscious states are conceived as mental states we are conscious *with*, making us conscious, as well as states we are conscious *of*, as they allegedly make us conscious *by* making us conscious of them (rather than by only making us conscious of the world), i.e., generating the inner awareness they are objects of. The purpose of Part I is to present the conceptions of consciousness involved in the articulation of these two alternative explanatory strategies, and to argue that the one presupposed by extrinsic higher-order theories offers significant theoretical advantages over the one presupposed by intrinsic higher-order theories. The first chapter focuses on the conception of consciousness presupposed by intrinsic higher-order theories and its intimate relation with the formulation of hard problem, while the second chapter offers a first reason why the conception of consciousness presupposed by extrinsic higher-order theories may be considered as being more fruitful than its alternative one: it is compatible with the rejection of the hard problem but it may also allow to address it (at least in part). Then, the third chapter will focus on the metaphysical implications of these two alternative conceptions of consciousness, and it will be argued that intrinsic higher-order theories presuppose a controversial metaphysical stance in the debate concerning the nature of fundamental properties (that is compatible with, but not entailed by extrinsic higher-order theories).

## 1. The Phenomenal Character View

The notion of phenomenal consciousness is notoriously hard to define, at least in any non-circular way (Block 1995). However, it is generally agreed that its reference can be fixed ostensively, by appealing to ‘what it is like’ for a subject to be in a conscious state (Nagel 1974). That is, a subject is phenomenally conscious iff there is something it is like to be that subject, and a mental state is phenomenally conscious iff there is something it is like for its subject to be in it (Chalmers 2004).

Theories of consciousness – trying to explain the existence and the nature of this ‘what it is like’ – are often framed in terms of phenomenal character,<sup>7</sup> i.e., the sum of the phenomenal properties of the mental states becoming conscious.<sup>8</sup> That is, consciousness is usually conceived as being primarily a property of mental states: there is nothing more to being a conscious subject other than having mental states that possess a phenomenal character. On this view, for example, there being something it is like for a subject to see the blue sky is for that subject to have a mental state with a bluish (subjectively given) qualitative character, to taste the sweetness of an apple is to have a mental state with a sweetish (subjectively given) qualitative character, and so on.

A popular justification for this conception of consciousness, endorsed by many philosophers advocating for very different positions across the logical space,<sup>9</sup> involves the thesis that phenomenal character is made of essentially conscious qualities whose instantiation is responsible for the constitution of consciousness.<sup>10</sup> This is *the phenomenal character view*: there is nothing more to being a conscious subject other than having

---

<sup>7</sup> The notion of phenomenal character can be used to refer to (a) the qualitative character of conscious states, (b) their subjective character, or ‘for-me-ness’, or (c) the compresence of (a) and (b). Option (b) will not be considered here, since it is rather obscure what it would mean to have a subjectively given conscious state without any qualitative character. For the sake of exposition, talks of ‘phenomenal qualities’ of conscious states will be used as concerning both options (a) and (c), i.e., whenever the two are not explicitly distinguished, what follows applies to conceptions of phenomenal character as ‘qualities with for-me-ness’ even when the ‘for-me-ness’ aspect of those qualities is not explicitly mentioned.

<sup>8</sup> The notion of phenomenal property can be defined as referring to individual aspects of one’s conscious experiences, or the properties that fix *what* it is like to be a subject, constituting the phenomenal contents of conscious states.

<sup>9</sup> For example, Block (1995), Chalmers (1996), Siewert (1998), Carruthers (2000), Levine (2001) and Kriegel (2009).

<sup>10</sup> The thesis that phenomenal character is made of essentially conscious qualities is not trivially true, because a state’s qualitativity may not depend on its being conscious. That is, it is not obvious that the qualitative aspects of conscious states must go out of existence whenever the mental states to which they are ascribed to are unconscious: the fact that phenomenal character is made of ‘experienced qualitative properties’ does not entail that those qualitative properties must be necessarily experienced (or ‘phenomenal’). For example, first-order representationalist theories of phenomenal qualities (e.g., Tye 1995) can certainly allow that one and the same quality (e.g., a certain abstract and nonconceptual intentional content) can occur both consciously and unconsciously. And even if phenomenal character is conceived as involving also subjective character, it is not obvious that the qualities that partly constitute the phenomenal contents of conscious experience cannot exist in an unconscious form when lacking ‘for-me-ness’.



mental states that possess a phenomenal character because consciousness is entirely constituted by phenomenal character itself. In other words, according to supporters of the phenomenal character view,

P-consciousness [i.e., phenomenal consciousness] is experience. P-consciousness properties are experiential ones. P-conscious states are experiential, that is, a state is P-conscious if it has experiential properties. The totality of *the experiential properties of a state are "what it is like"* to have it. Moving from synonyms to examples, we have P-conscious states when we see, hear, smell, taste, and have pains. P-conscious properties include the experiential properties of sensations, feelings, and perceptions, but I would also include thoughts, desires, and emotions (Block 1995, 230. Italics mine).

That is, on the phenomenal character view, there is nothing more to being a conscious subject other than having mental states that possess a phenomenal character because the notion of consciousness is equated to that of conscious experience (“P-consciousness is experience”) and the notion of experience is in turn construed in terms of a subject’s having mental states endowed with essentially conscious qualities (“P-conscious properties”) that constitute a state’s phenomenal character (“the totality of the experiential properties of a state”), thereby determining what it is like to be the subject of those states – hence, constituting phenomenal consciousness itself. For example, according to the supporter of the phenomenal character view, not only there being something it is like for a subject to see the blue sky is for that subject to have a mental state with a bluish phenomenal character but, moreover, it is that very bluish phenomenal character that, in virtue of its essential features, constitutes the subject’s phenomenal consciousness of it.

The purpose of this chapter is to present this explanatory strategy in more detail and to argue that it should not go unquestioned, insofar as it involves significant theoretical commitments. In what follows (§1.1), it will be argued that the phenomenal character view implies that consciousness is constituted by intrinsic qualities of conscious states, a thesis that in turn naturally leads to the formulation of the hard problem of consciousness. It will then be argued (§1.2) that because of these implications the assumption of the phenomenal character view not only precludes the possibility of adopting eliminativist and illusionist approaches to the hard problem but may also lead to misinterpret them. Finally, the chapter will be concluded by suggesting that, even while taking the hard problem at face value, the explanatory strategy embedded within the phenomenal

character view – adopted by the supporter of intrinsic higher-order theories – faces significant competition, insofar as taking the hard problem at face value does not directly entail conceiving consciousness as constituted by essentially conscious qualities.

### **1.1. Consciousness as Phenomenal Character**

The purpose of this section is to argue for the entailment from the phenomenal character view (i.e., the view that consciousness is constituted by essentially conscious qualities) to the conception of phenomenal qualities as intrinsic properties of conscious states, which directly leads (with the help of some plausible assumptions concerning the nature of scientific explanation) to the formulation of the hard problem. However, the defence of these two claims presupposes some clarificatory work first, insofar as conflicting uses of the notion of intrinsic are present in contemporary philosophy of mind. In particular, the claim that phenomenal properties are intrinsic qualities is sometimes interpreted as meaning that they are metaphysically unstructured entities, whose identity can be determined without reference to any kind of internal or external relation.<sup>11</sup> Were this interpretation of the notion of intrinsic presupposed, it would be trivial that the phenomenal character view does not entail the intrinsic nature of phenomenal properties, since phenomenal character can be conceived as an internal relation between the qualitative and the subjective character of conscious states. Thus, before considering the relationship between the phenomenal character view and the intrinsic conception of phenomenal properties, a weaker interpretation of the notion of ‘intrinsic’ will be defended. In what follows, it will be argued that the stronger interpretation of ‘intrinsic’ mentioned above is unduly theoretically loaded, insofar as no phenomenological intuition supporting the intrinsicness of phenomenal properties clearly entails that those properties’ identities must be determined without reference to any kind of internal relation.

#### ***1.1.1. Varieties of intrinsic qualities***

The distinction between intrinsic and extrinsic properties can be introduced by means of the contrast between something’s having some properties “in virtue of the way that thing itself, and nothing else, is” (Lewis 1983, 197) and having other properties in virtue of the way it is related to the world. Although this conception of intrinsic properties has

---

<sup>11</sup> For example, Rosenthal holds that “we would insist that being conscious is an intrinsic property of mental states only if we were convinced that it lacked articulated structure, and thus defied explanation” (1993b, 157).

been put to use in various works on consciousness and mind (e.g., Kim 1982, Horgan 1993, Jackson 1998), the notion of intrinsic has also been used in a more theoretically loaded sense: given the essentially relational nature of extrinsic properties, it may appear natural to treat the term ‘intrinsic’ as synonymous with the term ‘nonrelational’. But, according to Lewis’ definition, whether or not the two are really synonyms depends on how the notion of nonrelational is disambiguated: if ‘nonrelational’ is taken to mean ‘not made of relations with other entities’ then it turns out to be a synonym of ‘intrinsic’, if it is taken to mean ‘not being made of relations altogether’, it does not. That is, while the non-extrinsic nature of intrinsic properties trivially entails their being nonrelational in a weak sense – since intrinsic properties are, by definition, not made of relations with entities other than the entity they are ascribed to – being nonrelational in these terms is not obviously the same as being nonrelational in the stronger sense of ‘not being made of relations altogether’, as intrinsic properties might involve internal relations while being nonrelational in the non-extrinsic sense. For example, it seems that “my property of having an arm is a relational property, but it is not extrinsic”, since “I do not instantiate it in virtue of standing in a relation *to something that does not overlap with me*” (Kriegel 2009, 145). The difference between these two senses in which we may consider intrinsic properties to be nonrelational can be expressed in epistemic terms (1) or in metaphysical terms (2). A property may be said to be ‘comparatively intrinsic’ if nonrelational in the weak non-extrinsic sense, and ‘absolutely intrinsic’ if nonrelational in the stronger sense of not being made of relations altogether:

(1) P is an absolutely intrinsic property of X just in case P is an intrinsic property of X, and the proposition that X has P is not a priori derivable from R, a proposition that details all and only the . . . extrinsic properties of X’s parts. P is a comparatively intrinsic property of X just in case P is an intrinsic property of X, and the proposition that X has P is a priori derivable from R (Pereboom 2011, 94)

(2) P is an *absolutely intrinsic* property of X just in case P is an intrinsic property of X, and this instance of P is not necessitated by purely extrinsic property instances of parts of X.

By contrast,

P is a *comparatively intrinsic* property of X just in case P is an intrinsic property of X, and this instance of P is necessitated by purely extrinsic properties of parts of X (Pereboom 2014, 51).<sup>12</sup>

Given the distinction between comparatively and absolutely intrinsic properties, even granting that “there is certainly some intuitive basis to the idea that truths about consciousness concern nonrelational or intrinsic properties” one may doubt “whether they concern absolutely nonrelational or intrinsic properties”, as this second thesis appears to be “left open by anything we know or believe about consciousness, either as a matter of theory or as a matter of introspection” (Stoljar 2014, 25). The supporter of the absolutely nonrelational conception of phenomenal properties may reject the possibility that phenomenal properties are only comparatively intrinsic by reiterating traditional arguments in favour of the intrinsic conception of phenomenal properties – that are supposed to show how extrinsic facts about mental states do not exhaust all there is to know about consciousness (e.g., the knowledge argument and the argument from the conceivability of zombies) – by lumping together extrinsic and comparatively intrinsic properties in the set of mental properties we can know independently of consciousness:

So, consider Zombie-Stoljar. Zombie-Stoljar is just like Stoljar with respect to his extrinsic and comparatively intrinsic properties. But unlike Stoljar, he lacks consciousness. In particular, he feels nothing when he touches velvet. It seems conceivable that Zombie-Stoljar should exist. If it is, then what it is like for Stoljar to touch velvet cannot be a priori derived from propositions detailing all of his extrinsic and comparatively intrinsic properties. It follows that if his experiences involve an intrinsic property then they involve an absolutely intrinsic property. We could construct a similar argument using Mary, who knows all about the extrinsic and comparatively intrinsic properties of color vision before leaving the room and learns what this form of experience is like only when she leaves. Swapping the black-and-white room for a velvet-free room hardly weakens the intuition (Alter 2016, 803).

But it is doubtful that this line of argument can successfully establish the absolutely nonrelational nature of phenomenal properties. One obvious objection is that Zombie-Stoljar may be intuitively conceivable only for those philosophers who already interpret phenomenology as presenting us with absolutely nonrelational qualities. That is, if the

---

<sup>12</sup> In what follows, it will be suggested that, at least in the case of phenomenal properties, it is not clear that this metaphysical understanding of the distinction between absolutely and comparatively intrinsic properties can be properly translated in epistemic terms.

absolutely intrinsic conception of phenomenal properties is presupposed by the arguments mentioned above then, even though it may be left implicit, those arguments surely cannot be used to justify such a conception (unless one is ready to accept circular justifications). The supporter of the absolutely intrinsic characterization of phenomenal qualities might reply that, for example, “conceivability argument proponents do not argue in that way, at least not typically”, rather, “they tend to advance the premise that zombies are conceivable on direct, intuitive grounds without invoking absolute intrinsicness” (Alter 2016, 804). That is, the intuitions concerning phenomenal properties presupposed by the conceivability argument (and its neighbours) may be “epistemically prior” to considerations concerning the absolutely intrinsic nature of those properties – “in the sense that the intuitions are used to justify those considerations” – even if “the latter are explanatorily prior, in the sense that they provide a theoretical basis for the intuitions” (Alter 2016, 800).

Yet, the risk of circularity can only be avoided if it is possible to find an interpretation of those intuitions that allows one not to beg the question against comparatively intrinsic conceptions of phenomenal properties – and it is not clear that it is possible. On the one hand, the intuitive phenomenological observation that phenomenal qualities appear to involve something more than the extrinsic properties of conscious states is too weak to grant (e.g.) the conceivability of Zombie-Stoljar, since phenomenal qualities may be non-extrinsic in virtue of being comparatively intrinsic. Unless one has already question-beggingly presupposed that ‘intrinsic’ means absolutely nonrelational (instead of nonrelational in the non-extrinsic sense), it is impossible to establish the absolutely intrinsic nature of phenomenal properties on the basis of their non-extrinsic appearance alone (as the possibility of their being comparatively intrinsic is left open). On the other hand, even granting that we can conceive of phenomenal qualities as intuitively appearing nonrelational without merely relying on their non-extrinsic appearance (i.e., appearing such that, e.g., Zombie-Stoljar is conceivable), it is possible that we may not be able to distinguish, within phenomenology, between absolutely and comparatively intrinsic properties. In fact, this possibility seems to be left open by Pereboom’s definitions. For a property P of X may be comparatively intrinsic according to the metaphysical definition (2), i.e., necessitated by extrinsic properties of X’s parts, despite appearing as absolutely intrinsic according to the epistemic definition (1), i.e., such that the fact that X has P is

not a priori derivable from facts concerning the extrinsic properties of X's parts – whenever the necessary connection between P and the extrinsic properties of X's parts can only be discovered a posteriori. If that is the case, then intuitions to the effect that phenomenal properties appear in experience in such a way that, e.g., Zombie-Stoljar is conceivable do not seem to justify the conclusion that “some truths about consciousness do not concern only either extrinsic or comparatively intrinsic properties” (Alter 2016, 803), because those apparently absolutely intrinsic properties may turn out to be comparatively intrinsic instead. Therefore, it seems that the question of whether those properties are absolutely or comparatively intrinsic is indeed ultimately “left open by anything we know [...] about consciousness” (Stoljar 2014, 25). Accordingly, even though within the framework of the phenomenal character view the idea that phenomenal qualities are intrinsic properties is sometimes interpreted in absolutely rather than comparatively intrinsic terms, it seems preferable to remain neutral on the issue and go back to Lewis' (1983) definitions: intrinsic properties are nonrelational *qua* non-extrinsic, but they are possibly made of relations nonetheless. Once the notion of intrinsic is defined in these non-committal terms, allowing for relational (though non-extrinsic) conceptions of phenomenal character, it becomes possible to argue that, within the framework of the phenomenal character view, phenomenal qualities – *qua* essentially conscious – must be conceived as intrinsic properties of conscious states.

### ***1.1.2. Intrinsic qualities and the hard problem***

Within the framework of the phenomenal character view, consciousness is equated with conscious experience and experiences are taken to be entirely constituted by what is experienced, i.e., the qualitative properties of conscious states. Hence, if consciousness is to be conceived as being entirely constituted by the qualitative properties that make up the phenomenal contents of experience, those qualitative properties must be essentially conscious: otherwise, their instantiation could not be sufficient for the constitution of consciousness.

The qualities constituting the phenomenal contents of experience may be characterized as being essentially conscious in virtue of their (absolutely or comparatively) intrinsic features, or in virtue of their extrinsic relationship with entities other than the parts of the mental states they are ascribed to (e.g., properties of other mental states or properties of the subject). If phenomenal properties are made of intrinsically conscious qualities (i.e.,

of qualities that are essentially conscious in virtue of their own peculiar features alone), the phenomenal character view appears as intuitively plausible, since it is a small leap from taking those qualitative properties to be the distinctive intrinsic features of conscious states to taking them to constitute consciousness by themselves. But if the essentially conscious nature of those qualities depends on their extrinsic relationships, it is not clear how phenomenal character could be considered as entirely responsible for the constitution of consciousness; rather, consciousness would be naturally conceived as the property that makes those states or their contents conscious (i.e., that at least partly constitute their phenomenal characters), rather than as a property constituted by what is experienced (i.e., phenomenal character). This observation applies to both conceptions of phenomenal character – as qualitative character alone or as involving subjective character as well – though for different reasons.

If phenomenal character is taken to involve the presence of subjective character, or the ‘for-me-ness’ of what is experienced, as suggested by the intrinsic higher-order theorist, consciousness is characterized in terms of the conscious inner awareness of one’s own mental states: “to say that my experience has a subjective character is to point to a certain *awareness* I have of my experience [...] my conscious experience is not only *in me*, it is also *for me*” (Kriegel 2009, 8).<sup>13</sup> In turn, within the framework of the phenomenal character view, that relation of (inner) awareness is supposed to be constituted precisely by the presence of subjective character. That is, it is the for-me-ness of the qualitative properties of one’s mental states that allegedly makes the subject aware of those states and their qualities: “what makes a mental state phenomenally conscious at all (rather than a non-phenomenal state) is its subjective character” (Kriegel 2009, 10). Within this framework, the supporter of the phenomenal character view is committed to the claim that phenomenal properties are made of intrinsically conscious qualities because, if their subjective givenness were not conceived as an intrinsic feature of phenomenal properties, then it would not be possible to consider phenomenal character as the only entity responsible for the constitution of inner awareness. That is, the assumption of an extrinsic

---

<sup>13</sup> That is, on this view, consciousness does not only presuppose the subject as a metaphysical condition of possibility, due to the fact that conscious experiences cannot be free-floating independent entities but require an experiencing subject to whom something is phenomenally given: “to say that an experience is *for me* is precisely to say something more than that it is *in me*” (Kriegel & Zahavi 2016, 36). If phenomenal character involves subjective character, the subject is not only aware of what is phenomenally given in her experience, but also of the experience itself – as an aspect of the phenomenal contents of that experience.

characterization of subjective character directly leads to the idea that, rather than being the for-me-ness of conscious states that constitutes inner awareness, subjective character is the result of such a relation obtaining, i.e., that the presence of an extrinsic relation of inner awareness is already presupposed. Hence, the natural conclusion would be that the existence of phenomenal character depends on the presence of consciousness (defined as the relation of inner awareness holding between the subject and her own mental states) rather than being what constitutes consciousness (i.e., what establishes the relation of inner awareness).<sup>14</sup> Therefore, either the supporter of the phenomenal character view takes the qualities constituting the phenomenal contents of experience to be intrinsically conscious (by assuming that their for-me-ness is an intrinsic feature of those qualities), or the phenomenal character view becomes untenable: extrinsic characterizations of subjective character are incompatible with the thesis that consciousness is entirely constituted by the phenomenal properties of conscious states.

If phenomenal character is taken to involve the presence of qualitative character alone, the phenomenal contents of experience are characterized exclusively in terms of ‘outer’ awareness, rather than inner awareness of one’s own mental states. That is, the phenomenal contents of experience are not supposed to involve the awareness of the subjective givenness of conscious qualities, but only the awareness of those subjectively given qualities – which, within the framework of the phenomenal character view, are supposed to make the subject aware of themselves, thereby constituting consciousness. The supporter of the phenomenal character view endorsing this conception of phenomenal character is committed to the claim that phenomenal properties are made of intrinsically conscious qualities because, if their making the subject aware of themselves were not conceived as an intrinsic feature of those qualities, then phenomenal character could not be considered as the only entity responsible for the constitution of consciousness. That is, if the qualities constituting the phenomenal contents of experience were essentially conscious in virtue of their extrinsic relationship with other mental states or the subject, the existence of phenomenal character would depend on the presence of

---

<sup>14</sup> This is why Kriegel holds that, e.g., given the bluish phenomenal character of my visual perception of the sky, we should “think of the key feature of my conscious experience as *bluish-for-me-ness*”, i.e., characterizing the subjective character of that visual experience as intrinsic to its bluish phenomenal character.



some unconscious mechanism making those qualities conscious.<sup>15</sup> Hence, phenomenal character could not be conceived as being what constitutes consciousness alone. Thus, once again, either the supporter of the phenomenal character view takes the qualities constituting the phenomenal contents of experience to be intrinsically conscious or the phenomenal character view becomes untenable.

Taking stock, according to the phenomenal character view, there is nothing more to being a conscious subject other than having mental states that possess a phenomenal character because consciousness is entirely constituted by the essentially conscious qualities that make up the phenomenal character of conscious states. In turn, those qualities must be characterized as being essentially conscious in virtue of their (absolutely or comparatively) intrinsic features – unless one is ready to give up the thesis that consciousness is constituted by phenomenal character. For, if the essentially conscious nature of qualitative properties is explained in extrinsic terms, a theory of phenomenal character will not be an account of how those qualities generate consciousness – as suggested by the supporter of the phenomenal character view – but rather it will be an account of how their phenomenal nature is constituted by the instantiation of consciousness. Therefore, within the framework of the phenomenal character view, the qualities constituting phenomenal character must be characterized as being intrinsically conscious, i.e., as essentially conscious in virtue of their intrinsic features.

The phenomenal character view, by implying that phenomenal qualities are intrinsic properties, is naturally associated with the acknowledgement of the hard problem of consciousness, i.e., the recognition that the task of explaining why the physical phenomena responsible for the performance of cognitive and behavioural functions are accompanied by phenomenally conscious experience poses a peculiar and particularly difficult problem. For any other cognitive phenomenon (e.g., attentional mechanisms, rational control of behaviour, information integration, etc.), independently of how hard and complex it may be to come up with an explanation of that phenomenon, it is easy to understand the general strategy to follow – that is, finding a cognitive mechanism that

---

<sup>15</sup> For example, the extrinsic relation in virtue of which those qualities are poised for cognitive use (Tye 1995). Taking the relevant cognitive mechanisms to be conscious would mean giving up the thesis that phenomenal character is only made of qualitative character – e.g., if phenomenal character were supposed to involve the awareness of a mental state's qualities' poisedness for cognitive use, rather than awareness of those qualities alone, it would naturally be taken to involve the presence of subjective character as well.

can perform the relevant function – but it is not clear that this standard explanatory strategy can be applied to the case of conscious experience. The phenomenal character view (allegedly) explains why it is so: consciousness is constituted by intrinsic properties of conscious states, while the standard explanatory strategy “characterizes its basic entities relationally, in terms of their causal and other relations to other entities” (Chalmers 1996, 137), i.e., extrinsically. Hence, it follows that the standard explanatory strategy must be blind to the intrinsic features of mental states that, on the phenomenal character view, are supposed to constitute consciousness. Therefore, by implying that phenomenal qualities are intrinsic properties of conscious states, the phenomenal character view naturally leads to the formulation of the hard problem. In what follows, it will be argued that for the same reason the phenomenal character view not only precludes the possibility of adopting eliminativist (§1.2.1) and illusionist (§1.2.2) approaches to the hard problem but may also lead to misinterpret them. Finally, the chapter will be closed by suggesting that the explanatory strategy embedded within the phenomenal character view should not be taken for granted, even while taking the hard problem at face value (§1.2.3).

## **1.2. Approaches to the Hard Problem**

### ***1.2.1. Eliminativism:***

There is no meaningful hard problem to be solved.

*Eliminativism about consciousness.* The radical eliminativist rejects the very notion of consciousness and its ostensive characterization in terms of ‘what it is like’, usually on the grounds that our common-sense understanding of psychology is deeply mistaken (as suggested by, e.g., Dennett 1978). According to eliminativists about consciousness, the conscious/nonconscious distinction – at least when phrased in terms of ‘what it is like’ – simply does not cut mental reality at its joints, and thus it should be replaced with other concepts more faithful to the true nature of the mind (e.g., Churchland 1983). On this view, there is no meaningful hard problem to be solved because, once we have noticed that the concept of phenomenal consciousness does not refer, the hard problem ends up concerning a made-up property in need of no further explanation. Accordingly, eliminativism about consciousness can be described as the combination of a methodological thesis, i.e., that we should eliminate the concept of consciousness from

serious talks about the mind (because it is useless or even counterproductive), and its complementary ontological commitment, i.e., the view that we should eliminate phenomenal consciousness from the catalogue of existent entities.

However, although these two theses are often deeply related in the radical eliminativists' arguments, they do not straightforwardly entail each other. First, the ontological thesis does not imply the methodological thesis: *pace* Quine (1980), the claim that a theoretical construct has a non-existent referent does not entail that its use in science and philosophy cannot be useful: "an entity eliminativist might do away with some entity (e.g. atoms) but decide to preserve talk, thought, and practices associated with that entity in science [...] for their predictive and heuristic benefits" (Irvine & Sprevak 2020, 349).<sup>16</sup> Moreover, it is not even obvious that the methodological thesis implies the ontological thesis, since we could want to get rid of a concept simply because it is more misleading than useful (given certain purposes), rather than because the denoted object does not exist. On the one hand, the radical eliminativist claims that the conceptual distinction between conscious and unconscious mind must be totally eradicated from serious talks about the mind because it involves concepts depicting mental reality so inaccurately that we should take them to have non-existent referents. But, on the other hand, it is possible that the concept of phenomenal consciousness, while misrepresenting some aspects of our actual mental life (e.g., because of the theoretically significant assumptions commonly built into it, such as the intrinsic conception of phenomenal properties), also captures some essential aspects of its partly misrepresented referent. That is, one may agree with some of the motivations behind the eliminativists' methodological thesis – for example, one may believe that "discourse about conscious experience is too subjective, hard to verify, does not generalise well, does not pick out a natural kind, produces intractable disagreements" (Irvine & Sprevak 2020, 351) – but that same person, rather than wanting to eliminate the concept of consciousness altogether (i.e., accepting the eliminativists' ontological commitments), may try "to show that an alternative way of talking, thinking, and acting is available" and that "this proposed alternative discourse is, on balance, better for achieving our scientific goals than the one targeted for elimination" (Irvine & Sprevak

---

<sup>16</sup> For example, while constructive empiricists maintain an agnostic attitude towards the existence of atoms and instrumentalists explicitly deny their existence, they both recognize the usefulness of the concept of atom and atom-talks for scientific progress.

2020, 351). Perhaps mental states never have phenomenal characters (conceived as the sums of intrinsic qualities giving rise to the hard problem), but the concept of phenomenal consciousness may partly cut mental reality at its joints nonetheless – for it could only be partially misrepresenting an existent aspect of the mind. In that case, we may want to call for a revision of that concept, instead of its elimination. That is, by making the target of elimination a theoretically loaded word (such as ‘phenomenal character’) used to pin down a concept (i.e., consciousness), rather than the underlying concept itself, it is possible to weaken the radical eliminativists’ methodological thesis in such a way that it does not entail their ontological commitments. This approach is exemplified by another popular kind of eliminativist attitude towards the hard problem (i.e., the thesis that there is no meaningful hard problem to be solved), focused on questioning the intrinsic characterization of phenomenal properties, rather than the existence of consciousness itself.

*Eliminativism about phenomenal character.* Eliminating phenomenal consciousness is not the only possible way to claim that there is no hard problem to be solved: the same outcome can be obtained by eliminating phenomenal character – defined as what is constituted by the intrinsic qualities of conscious states – while remaining committed to the existence of consciousness and qualitative properties. That is, on this view, what must be eliminated is a (supposedly) wrong and theory-laden characterization of the phenomenon we want to explain, rather than the underlying phenomenon itself – e.g., we should accept that there is something it is like to see the blue sky, but reject that such ‘what it is like’ involves an intrinsic bluish quality ascribed to the mental state one is conscious of.

Since considerations in favour of the intrinsic conception of phenomenal qualities generally rely on how those qualities allegedly appear in experience, eliminativism about phenomenal character can be justified by questioning those phenomenological assumptions, i.e., by denying that phenomenal qualities subjectively appear as intrinsic properties. This move may be defended either by assuming that phenomenal properties appear extrinsic or by assuming that phenomenology does not imply theoretically substantive theses such as the intrinsic or the extrinsic nature of phenomenal properties. While the first option seems hardly defensible, the theoretical neutrality of phenomenology may be justified by questioning the popular assumption that, in

describing phenomenal properties, “the best you can do is use words to point to a phenomenon that the reader has to experience from the first person point of view” (Block 2015, 213). That is, the claim that consciousness does not involve the instantiation of (essentially conscious) intrinsic qualities may be defended by arguing that, if we want to give a description of consciousness that is not theory-laden, we should not rely on the subjective point of view of consciousness to fix the identity and existence conditions of the qualities we are conscious of – rather, we should only appeal to psychological phenomena that are not themselves conscious. For example, according to Rosenthal “exclusive reliance on subjective awareness is not a satisfactory way to approach our understanding of qualitative character” (2015, 37), because “the picture of qualities provided by our access via consciousness [...] tends to obscure the causal connections qualities stand in, promoting a conception of them as ineffable” (Coleman 2015, 21), and thus it is preferable to “describe and explain mental qualities not by appeal to what it’s like for us to be in conscious states that exhibit qualitative character, but instead by appeal to the role states with qualitative character play in perception” (Rosenthal 2015, 34).<sup>17</sup> By doing so, it becomes possible to doubt the intrinsic nature of the qualitative properties of conscious states (since they could be defined in terms of their causal profiles)<sup>18</sup>, and thus to question their essentially conscious nature: perhaps those qualities appear as essentially conscious only as long as we rely on the point of view of consciousness to establish what they are – because consciousness can only access them when they happen to be conscious.<sup>19</sup> Accordingly, eliminativism about phenomenal character turns out to be incompatible with the essential requirement of the phenomenal character view – namely, that the qualitative properties we experience in consciousness are intrinsically conscious.

Within the framework of the phenomenal character view, defending eliminativism about phenomenal character may appear as being only a superficially different way of

---

<sup>17</sup> In particular, Rosenthal holds that “the best way to take perceptual role into consideration will rely on the ability an individual has to discriminate stimuli”, so that mental qualities will be conceived as “the mental properties in virtue of which an individual can perform perceptual discriminations” (Rosenthal 2015, 34), and will thereby become classifiable in ‘quality-spaces’, empirically determined by each quality’s relative position on the basis of subjects’ responses to ‘just noticeable differences’ between distinct stimuli.

<sup>18</sup> It should be noticed, however, that it is also possible to define qualitative properties in extrinsic terms without denying that they have underlying intrinsic features nonetheless.

<sup>19</sup> The converse does not hold: one may doubt the essentially conscious nature of phenomenal properties without doubting their intrinsic nature. This fact is the foundation for the development of theories of consciousness involving the rejection of the phenomenal character view but maintaining the commitment to a realist attitude towards the hard problem.

arguing for the same substantial thesis defended by eliminativists about consciousness (i.e., that phenomenal consciousness should be removed from our ontology). For, on both types of eliminativism about the hard problem, nothing strictly speaking satisfies the concept of consciousness as defined within the framework of the phenomenal character view: despite remaining committed to the existence of consciousness and qualitative properties, the eliminativist about phenomenal character may not even appear, in the eyes of the supporter of the phenomenal character view, as trying to explain a different phenomenon while giving it the same name, i.e., consciousness (e.g. Block 1995). However, it is far from clear that this reconstruction of the debate is correct. Because, by avoiding the assumption of the phenomenal character view (i.e., the view that consciousness is constituted by essentially conscious intrinsic qualities), it becomes possible to assess the differences between these two conceptions of consciousness as having a theoretical, rather than a definitional nature. That is, by refusing to conceive consciousness as being entirely constituted by phenomenal character, it becomes possible to interpret the disagreement on whether phenomenal qualities are intrinsic or extrinsic as depending on the disagreement concerning the best way to determine their identity and existence conditions, rather than interpreting the latter disagreement as dependent on philosophers' targeting different explananda. Therefore, while the assumption of the phenomenal character view suggests that theories developed within the framework of eliminativism about phenomenal character are simply not viable accounts of consciousness (in that they are committed to the elimination of phenomenal consciousness), allowing that consciousness may not be entirely constituted by phenomenal character leads to a more neutral definition – compatible with conceptions of intrinsic as well as extrinsic conceptions of phenomenal properties, and allowing to genuinely discuss questions concerning how we should fix the identity and existence conditions of phenomenal qualities.

However, even if one construes both types of eliminativism about the hard problem as eliminating consciousness altogether, it is still possible to repurpose the weakened version of the eliminativists' methodological thesis in such a way that it does not entail their ontological commitments, even without rejecting the popular phenomenological assumption that phenomenal qualities subjectively appear as intrinsic properties. This is

the reason why an intermediate position between eliminativist and realist attitudes toward the hard problem exists.

### **1.2.2. Illusionism:**

There is an apparently hard problem, but it has an easy solution.

According to the illusionist, solving – or better, dissolving the problem of consciousness simply requires us to solve the ‘meta-problem’<sup>20</sup> (Chalmers 2018): once we have explained why it seems to us that we have conscious states, we have explained all there is to explain about consciousness (Frankish 2016).<sup>21</sup> While the radical eliminativist argues that the concept of phenomenal consciousness is (at best) irrelevant or useless, the illusionist does not need to deny the phenomenological datum that phenomenal consciousness is a genuine and important aspect of the mind that needs to be explained. Moreover, while the eliminativist about phenomenal character takes that phenomenological datum to be theoretically neutral (with respect to the nature of the qualities we are conscious of), the illusionist accepts the conception of phenomenal properties as being subjectively presented in experience as essentially conscious intrinsic properties. However, differently from the realist, the illusionist denies the existence of phenomenal character (defined as what is constituted by the intrinsic qualities of conscious states) and claims that the notion of ‘what it is like’ should be conceived in terms of conscious states *seeming* to have phenomenal character. For example, the illusionist accepts that there it is like to see the blue sky and that it subjectively seems that such an experience is constituted by an intrinsic bluish phenomenal character, but holds that this subjective appearance does not exist in virtue of those intrinsic features, rather, it is constituted by extrinsic mechanisms that make the resulting experience appear as if it were made of intrinsic properties. By pursuing this strategy, it becomes possible to dissolve the hard problem by explaining why it seems that there is a hard problem of consciousness to begin with – or why our conscious states seem to have peculiar qualitative properties. In other words, although the illusionist denies that what many would call phenomenal consciousness exists (just like eliminativists), illusionism cannot

---

<sup>20</sup> That is, the problem of explaining why it seems that consciousness poses a hard problem.

<sup>21</sup> The version of illusionism considered here is ‘weak illusionism’ (e.g., Graziano, 2016; Humphrey, 2016). For the alternative formulation of the view, i.e., ‘strong illusionism’, does not provide an alternative approach to the hard problem, but rather it assumes eliminativism about phenomenal character and tries to justify why many erroneously believe that eliminativism is incorrect by appealing to the shortcomings of introspection (e.g., Frankish 2016).

be interpreted by the supporter of the phenomenal character view as eliminativism about consciousness. For it does not entail that it is an illusion that consciousness exists, rather, it only requires the contents of conscious experiences to be essentially illusory. If consciousness is the illusion that conscious states have phenomenal characters, then it is not an illusion *that* consciousness exists: as long as conscious experiences are conceived as existent illusions we actually experience, the claim that consciousness is not constituted by intrinsic qualities of conscious states (as it allegedly seems to be) does not entail the claim that the illusion called consciousness does not exist. (For if it did not, there would be no meta-problem to address).

This weakened formulation of the eliminativist ontological thesis – involving the claim that there is in fact something we should call phenomenal consciousness although it is not made of the intrinsic properties it appears to be made of – may seem incompatible with the weakened methodological thesis mentioned earlier that, although the conception of consciousness as constituted by the intrinsic qualities of conscious states is erroneous, it may be useful nonetheless. But even though once again nothing strictly speaking satisfies the concept of phenomenal consciousness as defined within the framework of the phenomenal character view, the illusionist takes that concept to have a misrepresented existent referent, and attributes the responsibility of the shortcomings in our phenomenological descriptions (that the supporter of the phenomenal character view considers as revealing the nature of consciousness) to the illusory nature of consciousness – conceived as being constituted by a cognitive mechanism making experiences look as if there were indeed intrinsic phenomenal properties in the world. Hence, although the concept of phenomenal consciousness (*as* phenomenal character) is in fact considered as a somehow inadequate concept, since it supposedly does not represent mental reality accurately, it can be useful nonetheless, because it accurately portrays the object of the illusion that consciousness generates – leading to pose the question of why, within phenomenology, it looks like there are intrinsic qualia (even though there are none).

The weakened formulation of the eliminativist ontological thesis may be found problematic on its own, regardless of its ties with the methodological thesis. For the challenge of explaining how illusions of phenomenality can arise in non-phenomenal systems may appear to be just as hard as the challenge of explaining how phenomenal properties – as defined within the framework of the phenomenal character view – can



come into existence (Prinz 2016). That is, although illusionism is devised precisely to dissolve the hard problem, it can be interpreted as generating a similar and equally hard “illusion problem”: how can consciousness involve the intrinsic appearance of phenomenal properties without involving the instantiation of intrinsic properties?

This question certainly constitutes a significant challenge to the illusionist, but it does not obviously lead to conclude that illusionism is not a viable approach to the hard problem.<sup>22</sup> Yet, within the framework of the phenomenal character view, the illusion problem can be interpreted as implying the incoherence of illusionism: under the assumption that consciousness is entirely constituted by what is experienced the illusionist may be seen, by positing the illusion of phenomenality (involving the intrinsic appearance of phenomenal qualities), as also positing the existence of the very same intrinsic properties he attempts to eliminate. On the one hand, illusionism is distinguished from eliminativism about phenomenal character precisely because it is committed to the claim that consciousness, conceived as involving the subjective presentation of intrinsic qualities, is a real thing. After all, illusions are real things independently of whether they correctly represent the relevant reality: it is just what they are about (i.e., those apparently intrinsic qualities) that is not real (in the non-technical sense of illusion). But, on the other hand, if consciousness is entirely constituted by those apparently intrinsic qualities as suggested by the supporter of the phenomenal character view, then accepting the reality of such an appearance seems to imply that there are in fact intrinsic qualities constituting consciousness. These considerations may lead the illusionist into an impossible dilemma: if the intrinsic appearance of phenomenal properties is enough to constitute their reality, it seems that either there is no illusion because there are in fact intrinsic qualities (as the illusion of their existence is enough to make them real) or there is no illusion because there is no seeming that there are (as eliminativists about phenomenal character hold). However, by refusing to conceive consciousness as being entirely constituted by phenomenal character, it becomes possible to hold that the fact that (seemingly intrinsic) phenomenal appearances are real things does not entail that they provide a truthful perspective on what consciousness is: even though the intrinsic appearance of phenomenal character is conceived as a real thing, there may be some underlying reality

---

<sup>22</sup> That is, it is up to specific illusionist theories to suggest how to answer the question of how illusions of phenomenality can arise in non-phenomenal systems.

behind those (real) appearances which ultimately determines the nature of the qualities we are presented with by those appearances (independently of how those qualities are presented to us). That is, the illusionist (like the eliminativist about phenomenal character) may doubt the reliability of the subjective point of view of consciousness for determining the existence conditions of the qualities we are conscious of – while using that same point of view to ostensibly fix the reference of the explanandum (unlike the eliminativist about phenomenal character). And, by doing so, the illusionist may reconcile the intuition that phenomenal reality is a matter of phenomenal appearances (i.e., that appearances are real things) with the claim that the intrinsic appearance of phenomenal properties is not sufficient to constitute their reality (i.e., the claim needed to avoid the impossible dilemma involved in the illusion problem formulated within the framework of the phenomenal character view).

Clearly, even setting the phenomenal character view aside, it will still be possible to object that the challenge of explaining how illusions of phenomenality can arise in non-phenomenal systems is just as hard as the challenge of explaining how intrinsic phenomenal properties can come into existence. And until such a challenge is met, even though the illusionist can argue that his approach to the hard problem should be preferred in virtue of its more parsimonious ontology (Frankish 2016), it is difficult to deny that the introduction of intrinsic phenomenal properties into our ontology may help us come up with better theories of consciousness than illusionist ones. Thus, we are left with only one possible approach to the hard problem to consider – taking it at face value.

### ***1.2.3. Realism:***

The hard problem is indeed very hard.

As mentioned at the beginning of this chapter, phenomenal consciousness can be ascribed to subjects as well as mental states. Thus, any theory of consciousness must answer the following two questions:

(a) What is the nature of the qualitative properties that make up phenomenal character?

That is, the question of what kind of properties constitute the contents of conscious experience.

(b) What is the nature of phenomenally conscious subjects?

That is, the question of what kind of properties make a subject conscious.

According to the supporter of the phenomenal character view, we should only answer question (b) derivatively. If the existence of a conscious subject can be explained in terms of that subject's undergoing conscious states with a phenomenal character, and phenomenal character is in turn constituted by essentially conscious intrinsic qualities, then a theory of phenomenal character is all we need to get a complete theory of consciousness. However, this explanatory strategy should not go unquestioned, because to reject the phenomenal character view (thereby refusing to answer question (b) derivatively) does not mean to assume that the hard problem should not be taken at face value. According to the supporter of the phenomenal character view, the conscious relation between the experiencing subject and what is experienced is taken to be entirely constituted by what is experienced because consciousness is conceived as being an intrinsic property of conscious states – as it is supposed to be entirely constituted by their essentially conscious intrinsic qualities. But while the thesis that consciousness is intrinsic to conscious states (i.e., the phenomenal character view) implies that what is experienced in consciousness are in fact intrinsic qualities, the converse does not hold: accepting that consciousness involves the experience of intrinsic qualities does not imply assuming the phenomenal character view. Consciousness may be characterized as an extrinsic mechanism that is at least in part responsible for the conscious nature of the intrinsic qualities we experience: by refusing to answer the question of what kind of properties make a subject conscious solely in terms of one's answer to the question of what kind of properties are experienced in consciousness, it becomes possible to outline an explanatory strategy alternative to the phenomenal character view but nonetheless compatible with realism about the hard problem.

Taking stock, even though the formulation of eliminativist and illusionist approaches to the hard problem requires the rejection of the phenomenal character view, the rejection of the phenomenal character view does not entail neither eliminativism nor illusionism about the hard problem: conceiving consciousness as what constitutes experiences, rather than as what is constituted by what is experienced, does not preclude the possibility of characterizing what is experienced as intrinsic qualities of conscious states. The purpose of the following chapter will be to present the fundamental dimensions of variation involved in the articulation of this alternative explanatory strategy.

## 2. The Extrinsic View

The purpose of this chapter is to present the conception of consciousness presupposed by extrinsic higher-order theories (henceforth: the extrinsic view), according to which consciousness is a cognitive mechanism that is extrinsic to the mental states made conscious (that determines their being conscious), and to argue that although the adoption of the extrinsic view is often associated with eliminativism or illusionism about the hard problem, it may also prove useful to develop higher-order theories that are alternative to intrinsic theories but nonetheless compatible with a realist attitude towards the hard problem.

In the first section (§2.1), two fundamental types of extrinsic views compatible with realism about the hard problem will be introduced: the ‘modest’ extrinsic view, characterizing consciousness as the extrinsic property of conscious states making their qualities manifest (e.g., Coleman 2017), and the ‘ambitious’ extrinsic view, characterizing consciousness as the extrinsic property of conscious states constituting their phenomenal qualitativity. It will then be argued that these two types of extrinsic views may be developed by assuming an unorthodox conception of consciousness – as being primarily a property of subjects and only derivatively a property of mental states (henceforth: the subject view), rather than by sharing with the supporter of the phenomenal character view the idea that there is nothing more to being a conscious subject than having mental states with a phenomenal character (henceforth: the state view).<sup>23</sup> Then, the second section (§2.2) will be devoted to the defence of the legitimacy of conceiving the distinction between state and subject views as involving two substantial theoretical options (rather than as involving just a verbal difference), providing a viable alternative framework for understanding the nature of consciousness.

### 2.1. Varieties of Extrinsic Views

According to extrinsic views, consciousness is distinct from phenomenal character and at least in part responsible for its existence: the phenomenal qualities experienced in consciousness (or their alleged ‘for-me-ness’) do not constitute consciousness by themselves (as suggested by the supporter of the intrinsic view); rather, what we

---

<sup>23</sup> In particular, it will be argued that the state view can only allow the development of modest extrinsic higher-order theories, and that the adoption of the subject view is required to articulate ambitious extrinsic higher-order theories.

experience is made phenomenal by consciousness – conceived as a property extrinsic to the mental states made conscious. For example, according to the extrinsic theorist, there is something it is like for a subject to see the blue sky, rather than because that visual state has an intrinsic bluish phenomenal character, because the state is in a suitable consciousness-grounding relation (and possibly also because it had a bluish quality to be made conscious in the first place).

Two fundamental types of extrinsic views can be distinguished, by considering once again the two questions that theories of consciousness must address:

(a) What is the nature of the qualitative properties that make up phenomenal character?

That is, the question of what kinds of properties constitute the contents of conscious experience.

(b) What is the nature of phenomenally conscious subjects?

That is, the question of what kind of properties make a subject conscious.

Rejecting the phenomenal character view means refusing to answer question (b) derivatively, by appealing only to one's answer to question (a). But that can be done in two ways: by treating question (b) as being partly independent of question (a) or by trying to answer question (a) derivatively, in terms of one's answer to question (b). The first strategy is typically adopted by the eliminativist about phenomenal character,<sup>24</sup> according to whom the contents of conscious experience are constituted by extrinsic properties of first-order states – whose nature must be specified in one's answer to question (a) independently of one's answer to question (b) – and their becoming conscious does not depend on intrinsic features of those states.<sup>25</sup> The second strategy is typically adopted by the illusionist, according to whom we should answer question (a) in terms of one's answer to question (b), because an explanation of the apparently intrinsic properties that make up phenomenal character does not presuppose a theory of the underlying extrinsic properties

---

<sup>24</sup> Although strictly speaking the eliminativist about phenomenal character may be interpreted as suggesting that question (a) simply should not be answered – as there is no phenomenal character to explain – it still makes sense to ask him what kind of properties are experienced in consciousness, as opposed to what kind of properties make a subject conscious.

<sup>25</sup> The resulting explanatory strategy consists in answering question (b) only *partly* independently from one's answer to question (a) because although the properties that make a subject conscious are not strictly speaking the experienced properties of first-order states, the presence of the latter properties (at least as intentional objects) is still required in order to constitute conscious experiences (unless one accepts the possibility of conscious experiences *of nothing*).

of conscious states that are misrepresented as intrinsic in the illusion we call consciousness, but only a theory of the extrinsic cognitive mechanism responsible for the illusion. However, both strategies – i.e., answering question (b) partly independently from question (a), or deriving one’s answer to question (a) from one’s answer to question (a) – can also be adopted while taking the hard problem at face value. The first strategy yields ‘modest’ extrinsic theories of consciousness, according to which an account of the property making subjects conscious does not by itself count as a solution to the hard problem, whereas the second strategy yields ‘ambitious’ extrinsic theories of consciousness, according to which an explanation of what makes a subject conscious is also a complete account of what constitutes phenomenal character. The contrast between modest and ambitious extrinsic views can be presented in more detail as follows:

*Modest extrinsic view.* Consciousness is the extrinsic property of mental states that unveils their pre-existing qualitative aspects: phenomenal character is made of non-essentially conscious intrinsic qualities that (due to their intrinsic nature) can be rendered ‘phenomenal’ in virtue of the instantiation of consciousness.

*Ambitious extrinsic view.* Consciousness is the extrinsic property of mental states that constitutes their phenomenal qualitativity: phenomenal character is made of non-essentially conscious extrinsic properties of mental states that become intrinsic phenomenal qualities in virtue of the instantiation of consciousness.

In the case of the modest extrinsic view, solving the hard problem means providing independent answers to question (a), concerning the kind of properties that are experienced in consciousness, and question (b), concerning the kind of properties making a subject conscious. On the one hand, devising a theory of consciousness would require an answer to (b) that is independent of one’s answers to (a), because the intrinsic qualities of mental states are supposed to become phenomenal properties (i.e., part of *what* is experienced) only when their qualitativity is unveiled by consciousness. On the other hand, a theory of consciousness (conceived as the extrinsic property that makes subjects conscious) would not provide *ipso facto* a theory of phenomenal properties, because a complete explanation of what phenomenal properties are would also require an independent answer to question (a), i.e., an account of ‘proto-phenomenal properties’ –

those qualities that, in virtue of their intrinsic features, can be made conscious (allowing us to experience their qualitativity). This explanatory strategy is eminently exemplified by Lockwood's "disclosure view", according to which we should adopt "a kind of *naïve realism* with respect to phenomenal qualities", which should be conceived as "intrinsic attributes [of first-order states] *as* disclosed by awareness", and thus consciousness should be characterized "as kind of searchlight, sweeping around an inner landscape" and thereby "revealing qualities that were already part of the landscape, rather than [...] bringing these qualities into being" (Lockwood 1989, 163; see also Coleman 2022 for a recent defense of this view). Thus, for example, if there it is like for a subject to see the blue sky it is because the subject is in a visual state endowed with an intrinsic (possibly unconscious) bluish quality *and* at the same time is in another mental state in virtue of which she becomes able to experience that quality.

In the case of the ambitious extrinsic view, solving the hard problem means directly deriving an answer to question (a) – concerning the kind of properties that are experienced in consciousness – from one's answer to question (b) – concerning the kind of properties making a subject conscious. On this view, devising a theory of consciousness (conceived as the extrinsic property that makes subjects conscious) would provide *ipso facto* a theory of phenomenal properties – as consciousness would not only allow us to access the qualitative aspects of mental states, but rather it would be completely responsible for their phenomenal qualitativity. That is, going back to Lockwood's (1989) metaphorical characterization of consciousness, the ambitious extrinsic theorist may claim that, in the same way in which the colours of external objects are constituted by the reflection of the light off their surfaces (whose reflectance profiles can be fixed extrinsically), so too phenomenal qualities are constituted by the interaction of one's 'inner searchlight' with the extrinsically determined contents of first-order states. The natural outcome of this view is the idea – reminiscent of the intrinsic higher-order theorist's conception of subjective character as the feature of phenomenal character that fixes its existence conditions – that "the core of the hard problem is posed not by the qualities themselves but by our experience of these qualities: roughly, the distinctive phenomenal way in which we represent the qualities or are conscious of them" (Chalmers 2018, 30). Clearly, the supporter of the ambitious extrinsic view cannot share with the supporter of intrinsic higher-order theories the further claim that the subjective character of conscious states

(allegedly responsible for the constitution of mental states' phenomenal qualitativity) is the intrinsic relation between their 'for-me-ness' and their experienced contents (e.g., Kriegel 2009, 11), unless one is ready to abandon the core principle of the extrinsic view, i.e., that consciousness is not part of the phenomenal character of the mental state made conscious. Thus, the viability of the ambitious extrinsic view depends on whether an extrinsic property of a mental state could become an intrinsic phenomenal quality by acquiring an extrinsic subjective character, i.e., a kind of 'for-me-ness' that involves a relation between the subjectively given qualities of the mental state made conscious and an entity distinct from that mental state (instead of a relation that is intrinsic to its phenomenal character). That is, the prospects of the ambitious extrinsic view depend on whether, for example, the extrinsically definable represented blueness of the sky may lack qualitative features when unconscious and become a bluish phenomenal quality, intrinsic to the subject's conscious experience (though not to the first-order state itself), in virtue of being related with its subject through his inner awareness.

In turn, this feature of the ambitious extrinsic view makes it incompatible with the conventional framework usually shared by intrinsic and extrinsic higher-order theories: *the state view*, i.e., the conception of consciousness as being primarily a property of mental states, such that there is nothing more to being a conscious subject other than having mental states that possess a phenomenal character. The core commitment of the state view can be further spelt out by assuming the common definition of mental states in terms of property-exemplification, i.e., as the instantiation of mental properties by an entity *S* at a time *t*. Thus, the conception of consciousness as being primarily a property of mental states can be described as a view that ascribes explanatory and ontological priority not to the subject's property of consciously experiencing certain phenomenal qualities, but rather to the property of the consciously experienced properties of giving rise to that experiencing. Examples of the widespread commitment to this view can be found in the work of intrinsic as well as extrinsic higher-order theorists.<sup>26</sup> On the one hand, the state view is directly implied by intrinsic higher-order theories, according to which the relation of (conscious) inner awareness between subject and conscious states is constituted by intrinsic properties of the latter. Thus, for example, Kriegel suggests that

---

<sup>26</sup> Other exemplary cases of commitment to the state view include Block (1995), Tye (1995), Chalmers (1996), Crane (2001), Levine (2001), Hill (2009) et al.



when considering “a specific, preferably simple, moment or episode of conscious experience, such as seeing blue or tasting chocolate or feeling nervous about an upcoming public lecture [...] the object of our contemplation is a property of states, not creatures” (2009, 32), and holds that it is such a property of states that must be explained in order to solve the mystery of consciousness, insofar as (e.g.) “the key feature of my conscious experience” of the blue sky – responsible for my experiencing of it – is taken to be the “*bluish-for-me-ness*” (2009, 11) ascribed to the experience itself. That is, according to Kriegel, it is not the case that there is something it is like for a subject to see the blue sky because the subject has the property of experiencing the contents of that visual state; rather, the subject’s property of experiencing is reduced to the subject’s property of having a mental state (or certain mental properties at *t*) whose intentional content is endowed with (bluish) *for-me-ness*.<sup>27</sup> On the other hand, the state view is also typically adopted by extrinsic higher-order theorists, who often take the property of subjects of being in a conscious first-order state (or experiencing it) to be reducible to the property of being in a mental state (or having certain mental properties at *t*) whose intentional content has the extrinsic property of being the intentional object of a suitable higher-order state. For example, Lycan explicitly distinguishes between “Q-properties”, defined as “introspectible (apparently) monadic qualitative properties inhering in a mental state, such as the color occupying such-and-such a region of your ordinary visual field right now” and “what it is like for the subject *to experience* a particular quale (in the first sense [i.e., Q-property])”, and argues that, e.g., “subjective redness is a Q-property, but the higher-order ‘what it’s like’ is a property of that Q-property itself” (2008, 237-8). Thus, despite their differences, Lycan (together with most extrinsic theorists) agrees with Kriegel that the subject’s property of consciously experiencing certain phenomenal properties – there being something it is like to have them – can be reduced to a property of mental states, i.e., that the conscious nature of the subject is due to a property (or a “feature”) of the experienced properties, rather than to a property of the subject that allows him to experience them – or even to constitute their phenomenal qualitativity, if the ambitious extrinsic view can be defended.

---

<sup>27</sup> Even though Kriegel explicitly avoids talks of properties of properties, by arguing that “subjective character and qualitative character are not separate properties, but in some (admittedly problematic) sense are aspects of a single property” (2009, 11), his upshot is still clearly to ascribe explanatory and ontological priority to a feature of the consciously experienced properties, rather than to a feature of the subject’s property of consciously experiencing them.

Extrinsic state views are incompatible with the ambitious extrinsic view because of their necessary commitment to the non-existence of intrinsic subjective character: since the experienced qualitative properties of mental states are supposed to be made phenomenal by inner awareness, which is characterized as a relation between two mental states, inner awareness itself cannot be conceived as a conscious property without giving rise to a familiar infinite regress of conscious states.<sup>28</sup> But rather than simply dismissing the ambitious extrinsic view as incoherent, it may prove useful to consider the consequences of rejecting the state view.

Even though consciousness may be characterized as an extrinsic property while still being considered as primarily a property of mental states – if it is conceived as an unconscious relation between mental states making their subject conscious – it may also be taken to be an extrinsic property of mental states because it is an intrinsic property of the subject that establishes a conscious relation with her states (making them conscious instead of being made conscious by them). It seems that, in principle, the fundamental tenet of extrinsic views – that the properties we experience are made phenomenal by consciousness – does not directly imply that both relata in that extrinsic relation should be conceived as mental states: consciousness may be primarily a property of the subject, rather than being primarily the property of one’s mental states (making one conscious). Call this *the subject view*. The basic idea behind the subject view is that we should reverse the explanatory strategy embedded within the state view. Instead of trying to explain consciousness by characterizing conscious subjects derivatively, as entities with mental states endowed with phenomenal characters, conscious states should be conceived as those mental states that are made conscious by a conscious property of their subject:

On this picture, it is *not* that the experience itself has some phenomenal properties, and then the subject is in some way put into contact with that conscious experience (including its phenomenal properties) and therefore comes to be in a certain conscious state. [...] It is not the conscious experience itself that is explaining why the subject is in a particular conscious state (Taylor 2020, 3499).<sup>29</sup>

---

<sup>28</sup> Questions concerning this regress will be considered in Part II (§4).

<sup>29</sup> This point should not be interpreted as only concerning the phenomenal character view just because on extrinsic views the experience acquires phenomenal property *in virtue of* being put into contact with the subject. For, if it is a mental state’s being related to an unconscious higher-order state that makes the subject aware of its content, it follows

The idea that consciousness should be considered as being primarily a property of subjects has been recently defended by arguing that the state view should be rejected because talk of phenomenal qualities of mental states is simply the result of an erroneous description of conscious experience:

Experiences have ‘qualitative character’ only in the sense that they involve an experiencing subject who instantiates experiential properties—properties such that there is something that it is like to have them. [...]

Experiences are a subclass of events. Events may be understood as involving things which instantiate properties. Subclasses of events can be distinguished by the kind of individuals involved in the event and by the kind of properties they instantiate. The subclass of experiences can be characterized by saying that the individuals involved are experiencing subjects who instantiate experiential properties. For instance, your experience of blue in a given moment consists of you (an experiencing subject) instantiating the experiential property of being phenomenally presented with blue (Nida-Rümelin 2018, 3361-2).

When one has an experience, one is modified in a certain way: one comes to have a certain property. One instantiates a property (e.g. the property of tasting a lemon) over a certain time, and this instantiation of the phenomenal property over the time constitutes a conscious experience [...] phenomenal properties are properties instantiated by subjects, they are not instantiated by experiences themselves. [...]

Phenomenal properties partially *constitute* conscious experiences but such properties are not *instantiated* by these experiences. Phenomenal properties do not modify or characterise the experience itself. Rather, they modify the subject in a certain way, and this modification of the subject by a phenomenal property constitutes the conscious experience (Taylor 2020, 3498-9).

However, it is doubtful that defending the idea that consciousness is primarily a property of subjects requires the ascription of phenomenal qualities to subjects themselves, rather than mental states, as suggested in the passages above. According to these proposals, conscious experiences should be conceived as events that are in turn characterized in terms of property-exemplification, i.e., in terms of a substance  $s^{30}$  having

---

that consciousness is supposed to be first and foremost a relation between mental states rather than being primarily a property of the subject of those states.

<sup>30</sup> For example, “things like tables, chairs, atoms, living creatures” (Kim 1976, p. 33).

a property *P* at a time *t* (Kim 1973; 1976).<sup>31</sup> Then, the existence of phenomenal qualities of mental states is denied by identifying the substance *s* with the experiencing subject and the property *P* with the property of being phenomenally given with certain phenomenal qualities. Yet, it is not clear why the property *P* should not be analysed in terms of the property of having mental states endowed with phenomenal qualities.<sup>32</sup> Taylor suggests that the conception of experience under consideration is appealing because of its ability to account for two plausible intuitions: that “for each particular experience, that very experience could not have been had by any subject other than the one that has it” and that “there is no experience such that a particular subject *must* always have it” (2020, 3494). But, unless one assumes the phenomenal character view (according to which consciousness is constituted by phenomenal character), there is no apparent reason to consider these intuitions as incompatible with the idea that mental states have phenomenal qualities – defined in turn as individual aspects of one’s conscious experiences, or as the properties that fix what it is like to be a subject. That is, adopting the explanatory strategy according to which “it is *not* that the experience itself has some phenomenal properties, and then the subject is in some way put into contact with that conscious experience” (Taylor 2020, 3499) does not preclude the possibility of accepting the existence of phenomenal properties of mental states – unless one presupposes that those properties are entirely responsible for the existence of the conscious experience itself (rather than only for fixing its identity). In other words, the fact that a mental state’s having phenomenal qualities is not what explains *why* consciousness exists does not entail that the properties of that mental state (e.g., what it is about) do not determine what is experienced in consciousness, i.e., “the ‘*what*’ it’s like for me” (Levine 2001, 7) to have a conscious experience.

---

<sup>31</sup> For an alternative characterization of events, in terms of particulars that are individuated by causal profile, see Davidson (1969; 1970). Examples of application of this view can be found in Tye (1995), Steward (1997), and Crane (2001). Taylor rejects this alternative view on the grounds that it makes it impossible to understand the plausible idea that a token experience has a certain subject necessarily, i.e., that it is not clear how “*other* subjects can have as intimate a relationship to a conscious experience as the subject who in fact has it” (2020, 3502).

<sup>32</sup> Following Steward (1997), it may be argued that the property *P* *should* be analysed as being the property of having mental states endowed with phenomenal qualities, because otherwise it would appear that experiences are not “entities about whose nature and properties questions can intelligibly arise which are not simply dependent on answers to prior questions about relations between other entities” (1997, 31). However, it will be argued in what follows that – even granting that experienced qualities are properties of first-order mental states – it is doubtful that the state view it is trivially true, since the relevant relations may directly involve the subject, rather than only the subject’s mental states.

Similar considerations apply to Nida-Rümelin's defense of the radical subject view, which is explicitly formulated as only arguing against the idea that mental states have 'qualia' conceived as essentially conscious properties that make the subject conscious of them (2018, 3364-6). She dismisses the relevance of a weaker interpretation of the notion of phenomenal qualities of mental states by insisting that properties such as "colours, sounds, tastes or odours [...] are not properties of experiences, rather they belong to what is present to the subject who undergoes the experience; they belong to its content" (2018, 3367). But it is not clear why the contents of experience should not be conceived as being properties of the mental states made conscious (i.e., as properties of experiences), rather than as properties directly ascribed to the subject (i.e., as "experiential properties" (2018 3362)). Ultimately, what matters for considering consciousness as being primarily a property of subjects is not to deny that phenomenal properties of mental states are involved in the constitution of consciousness, but to hold that those properties of mental states can only determine *what* it is like to undergo a given experience, instead of determining *that* there is something it is like to be an experiencing subject.

Therefore, it seems that the generally agreed-upon thesis that consciousness involves the instantiation of phenomenal qualities by mental states does not entail that the subject comes to be in a certain conscious state by being "in some way put into contact with that conscious experience" and its phenomenal qualities (Taylor 2020, 3499). For the phenomenal nature of those qualities may be determined by a property of the subject, establishing a conscious relation with the mental states instantiating them – i.e., endowing those states with extrinsic subjective givenness and thereby making those states and their extrinsic (proto-)qualities conscious.

Accordingly, the subject view may be developed in less radical forms: rather than ascribing phenomenal properties to subjects instead of mental states, it is possible to ascribe consciousness to subjects – conceived as the property responsible for the phenomenal givenness necessary for the constitution of phenomenal qualities – and phenomenal characters (made of sums of phenomenal qualities) to mental states – responsible for fixing the identity of what is phenomenally given.

In what follows (§2.2), this unorthodox explanatory strategy will be further developed by appealing to the conceptual distinction between ascriptions of phenomenal

consciousness to subjects and ascriptions of phenomenal consciousness to mental states considered earlier,<sup>33</sup> in order to suggest the possibility of drawing a distinction between two different kinds of phenomenal properties – phenomenal perspectives,<sup>34</sup> denoted by ascriptions of phenomenal properties to subjects, as opposed to phenomenal characters, denoted by ascriptions of phenomenal properties to mental states. In turn, talk of phenomenal perspectives will help in understanding the way(s) in which mental states could be made conscious in virtue of the conscious nature of their subject:

*The subject view.* Consciousness is primarily a property of subjects because mental states become conscious in virtue of being caught up within their subject's phenomenal perspective.

The subject view defined in this (non-radical) manner provides a viable framework (alternative to the state view) for the articulation of extrinsic views of consciousness in both their modest and ambitious formulations – insofar as phenomenal perspectives may be conceived as equally ontologically fundamental for the constitution of conscious experience as the intrinsic qualities of our mental states, or as entirely responsible for the constitution of the phenomenal qualitativity of conscious states:

*Modest subject view.* Consciousness is the property of subjects that unveils the intrinsic qualities of mental states, thereby transforming them into phenomenal properties: the contents of experience (i.e., phenomenal character), are made of non-essentially conscious intrinsic qualities that (due to their intrinsic features) can be rendered 'phenomenal' in virtue of being caught up within the subject's phenomenal perspective.

*Ambitious subject view.* Consciousness is the property of subjects that constitutes the phenomenal qualitativity of mental states: the contents of experience (i.e., phenomenal character) are made of non-essentially conscious (intrinsic or extrinsic) properties of mental states that undergo intrinsic

---

<sup>33</sup> That is, the conceptual distinction justifying the difference between the two fundamental questions any theories of consciousness should answer (namely, (a) what kind of properties are experienced in consciousness, and (b) what kind of properties make a subject conscious).

<sup>34</sup> The notion of phenomenal perspective will be better specified later (§2.2), but it can be provisionally characterized in metaphorical terms as referring to the property in virtue of which an entity can occupy "the [conscious] standpoint or the position of a *person* or *subject*: the 'place' from which they 'see' things" (Crane 2001, 4).

modifications (e.g., acquire the intrinsic qualitative features we experience) in virtue of being caught up within the subject's phenomenal perspective.

The remainder of this chapter will be devoted to the defense of the viability of the subject view against possible objections found in the literature. After arguing in favour of the legitimacy of drawing a metaphysical distinction between phenomenal characters and phenomenal perspectives, it will be argued that the adoption of the subject view may provide the supporter of extrinsic higher-order views with powerful explanatory tools unavailable to the supporter of the state view – hence, that the subject view deserves more consideration than it has received so far in the contemporary discussion.

## **2.2. Varieties of Phenomenal Properties**

Phenomenal consciousness involves the instantiation of phenomenal properties: if there is something it is like to be a subject then certain phenomenal properties are instantiated by that subject, and if there is something it is like to be in a mental state certain phenomenal properties are instantiated by that state (Chalmers 2004). Ascriptions of phenomenal properties to subjects and to mental states denote conceptually distinct kinds of properties: in calling a mental state 'conscious', we are ascribing to that state some experienced qualitative properties; in calling a subject 'conscious', we are ascribing to that subject the property of consciously experiencing those qualitative properties.<sup>35</sup> This conceptual difference can be framed in familiar terms by appealing to the distinction between 'state-consciousness', i.e., the property distinctive of phenomenally conscious mental states, and 'creature-consciousness', i.e., the kind of consciousness we ascribe to subjects rather than mental states.

This distinction, introduced by Rosenthal (1986), is sometimes used to endorse the state view, i.e., to hold that we can ascribe phenomenal properties to subjects only in virtue of the fact that those subjects have mental states with phenomenal character (Carruthers 2000; Rosenthal 2005; Gennaro 2012). For 'creature-consciousness' was initially defined by Rosenthal as follows: "for an organism to be conscious means only that it is awake, and mentally responsive to sensory stimuli" (1986, 351). Given this

---

<sup>35</sup> For the philosopher sympathizing with the idea that consciousness is diachronically unified, the same point may be rephrased as follows: in calling a mental state 'conscious', we are ascribing to that state some qualitative properties that appear in the subject's stream in consciousness; in calling a subject 'conscious', we are ascribing to that subject a stream of consciousness, which may or may not be reducible to the qualities that appear in it (or to the states instantiating those qualities).

definition, it is apparent that there may be conscious creatures who do not (or cannot) instantiate phenomenally conscious states, since neither wakefulness nor responsiveness to stimuli can be conceived as co-extensive with phenomenal consciousness (e.g., Damasio 1999, §3). Hence, given this definition of creature-consciousness, it is natural to conclude that “what it is like to be a particular conscious individual is a matter of the sensory qualities of that individual's conscious experiences” (Rosenthal 1986, 352), i.e., that the state view is true: subjects do not have consciousness in virtue of their being conscious creatures, but only insofar as their mental states have state-consciousness. Yet, by taking the notion of creature-consciousness to denote wakefulness alone, Rosenthal et al. do not provide any reason to believe that the state view is correct – they just assume so, overlooking the subject view without providing substantial justification. Even if wakefulness can be legitimately considered as a kind of creature-consciousness, since it is a property of subjects rather than mental states, it is not necessarily the only kind of creature-consciousness there is (Bayne 2007, 14; Kriegel 2009, 26). Thus, the distinction between state-consciousness and creature-consciousness may be reformulated in a more neutral manner – that does not lead to overlooking the possibility of conceiving phenomenal consciousness as being primarily a property of subjects, rather than being primarily a property of mental states:

*State-consciousness.* A mental state M is conscious iff M has a phenomenal character, i.e., phenomenal properties determining what it is like to be in M.

*Creature-consciousness.* A subject S is conscious iff S has a phenomenal perspective, i.e., phenomenal properties determining that there is something it is like to be S.

According to the supporter of the state view, creature-consciousness can be wholly explained in terms state-consciousness: subjects can be said to have a phenomenal perspective simply in virtue of having mental states that possess a phenomenal character – which may be in turn constituted by intrinsically conscious qualities (as suggested by the supporter of the phenomenal character view) or at least in part constituted by extrinsic relations with other mental states (as suggested by the supporter of extrinsic views). According to the supporter of the subject view, creature-consciousness cannot be explained solely in terms of state-consciousness: the phenomenal properties constituting



phenomenal perspectives are not just the phenomenal properties of the mental states made conscious – which may be in turn partly constituted by the intrinsic qualities of mental states (as suggested by the supporter of the modest subject view) or wholly constituted by mental states’ being caught up within a phenomenal perspective (as suggested by the supporter of the ambitious subject view).

The *prima facie* plausibility of the idea that state-consciousness may not be ontologically more fundamental than creature-consciousness rests on the observation that the notion of phenomenal perspective is epistemically prior to the notion of phenomenal character: we can understand the notion of phenomenal character only once we grasp what it is like to be a subject, i.e., what it means to have a phenomenal perspective (Paternoster 2014, 251; Nida-Rümelin 2018, 3363). That is, while the notion of ‘what it is like to be a subject’ relies on the intuitive contrast between a property that we have and that inanimate objects such as tables and chairs lack (i.e., consciousness), no such contrast is available in the case of mental states unless we already assume the subjective point of view of consciousness (Speaks 2015, 3), within which we can describe conscious states as having properties that unconscious states lack. In fact, when Nagel put forward the idea that consciousness should be conceived in terms of ‘what it is like’, he introduced the notion of conscious state derivatively, by appealing to there being something it is like to be a conscious subject, rather than to the phenomenal properties of mental states:

An organism has conscious mental states if and only if there is something that it is like to be that organism – something it is like *for* the organism (Nagel 1974, 436).

The supporter of the state view may (rightfully) object that this kind of epistemic priority of phenomenal perspectives does not entail that state-consciousness is not ontologically more fundamental than creature-consciousness. For, even granting that the notion of phenomenal character is a theoretical term introduced from the subjective point of view of consciousness by relying on our understanding of what it means to have a phenomenal perspective – so that we have no independent grip on the kind of property phenomenal character refers to – it may be the case that, nonetheless, it is the notion of phenomenal character that picks out the fundamental ground of phenomenal consciousness. However, the purpose of this section is not to deny this possibility (i.e.,

that the state view *may* be correct), but only to defend the claim that the subject view can provide a viable alternative explanatory strategy (with its own theoretical advantages).

The significance of the distinction between phenomenal character and phenomenal perspectives is usually overlooked not because of convincing arguments in favour of the ontological priority of state-consciousness, but only because creature-consciousness can be defined derivatively in terms of having mental states endowed with phenomenal characters.<sup>36</sup> For example, according to Chalmers, “it does not make much difference whether one focuses on the phenomenal properties of subjects or of mental states” because “it is easy to translate between the two ways of talking” (2004, 155). Similarly, Kriegel holds that there is nothing controversial in theorizing about consciousness by assuming the framework of the state view because the notion of creature-consciousness can be analysed in terms of state-consciousness, since no subject can be phenomenally conscious without being capable of instantiating phenomenally conscious mental states:

A creature C is [...] creature-conscious iff there is a mental state M, such that (i) M is [...] state-conscious and (ii) nomologically possibly, C is in M (2009, 29).

Clearly, a biconditional of this kind cannot be used to defend the ontological priority of phenomenal characters over phenomenal perspectives, as it also allows to define state-consciousness in terms of creature-consciousness: even if a conscious subject must (nomologically possibly) have conscious states, it does not follow that she is phenomenally conscious *in virtue of* the fact that some of her mental states instantiate certain phenomenal properties – i.e., that the existence of consciousness is to be explained only by appealing to the phenomenal characters of those states. For, just as one may try to define the instantiation of phenomenal properties by subjects derivatively, in terms of the instantiation of mental states endowed with phenomenal characters, the instantiation of phenomenal properties by mental states may be defined derivatively in terms of those states being caught up within their subject’s phenomenal perspective.<sup>37</sup> Thus, on the one hand, given that the interdefinability of state-consciousness and creature-consciousness does not entail the ontological priority of state-consciousness, the supporter of the state

---

<sup>36</sup> E.g., Block (1995), Tye (1995), Chalmers (1996), Crane (2001), Levine (2001), Kriegel (2009), Hill (2009) et al.

<sup>37</sup> For example, Kriegel’s biconditional may be reversed as follows:

A mental state M is state-conscious iff there is a subject S, such that (i) S is creature-conscious, (ii) S instantiates M, and (iii) M is part of the phenomenal perspective of S.

view can only take this datum to imply that the distinction between phenomenal character and phenomenal perspectives rests on a conceptual difference (between ways of ascribing consciousness) that does not necessarily reflect a deeper metaphysical difference (between kinds of phenomenal properties). Yet, on the other hand, the validity of the inference from the interdefinability of state-consciousness and creature-consciousness to the claim that we should treat the notions of phenomenal character and phenomenal perspective as being ontologically equivalent is far from being trivial, either.

State-consciousness and creature-consciousness may be interdefinable, not because they are made of one and the same property (i.e., consciousness conceived as the property of mental states of having a phenomenal character), but because they are made of distinct properties that are essentially related: the fact that a subject's having creature-consciousness necessarily involves the nomological possibility of that subject's having state-conscious mental states does not imply that the property distinctive of conscious states (i.e., phenomenal character) exhausts the nature of consciousness – such that, e.g., “what it is like to be Yvette at *t* is constituted by what it is like for Yvette to be in the maximal [i.e. total] conscious state she is in at *t*” (Kriegel 2009, 31, fn. 17).

That is, talk of phenomenal characters and phenomenal perspectives may refer to metaphysically distinct kinds of phenomenal properties that are both necessary for the constitution of conscious experience – the former determining the identity of specific experiences (i.e., their phenomenal contents), and the latter determining their existence conditions (i.e., their being conscious). In that case, one kind of phenomenal properties could ground the other (as suggested by the ambitious subject view, according to which a mental state's qualitativity is due to that state's being caught up within the subject phenomenal perspective), or these two kinds of phenomenal properties may exist partly independently of each other while being only jointly sufficient for the constitution of conscious experience (as suggested by the modest subject view, according to which mental states can acquire phenomenal character when caught up within the subject phenomenal perspective only because they already possess intrinsic ‘proto-phenomenal’ qualities).

And, given these possibilities, it seems that the claim that being a conscious subject (i.e., having a phenomenal perspective) consists in nothing more than having mental

states with phenomenal characters cannot be justified solely on the basis of the interdefinability of state-consciousness and creature-consciousness without begging the question against the subject view. That is, taking the state view to be an uncontroversial framework only because the notion of creature-consciousness can be analysed in terms of state-consciousness simply means hiding one's (nontrivial) ontological commitment to the claim that creature-consciousness is not a 'substantial' entity, made of a kind of phenomenal property that manifests itself in conscious experience but is not constituted by the property of mental states of having a phenomenal character.

The supporter of the state view may object that the subject view inflates our ontology unnecessarily, and thus that considerations of ontological simplicity strongly suggest that the state view should be preferred. Yet, while this objection may have some plausibility in the eyes of the supporter of extrinsic states views, it is significantly less convincing once the notion of conscious subjective character is introduced. For, while the supporter of intrinsic higher-order theories must find a way of defending the idea "that subjective character and qualitative character are not separate properties, but in some (admittedly problematic) sense are aspects of a single property" (Kriegel 2009, 11), the introduction of phenomenal perspectives into our ontology frees the higher-order theorist from the need to find such an explanation (by conceiving subjective character as extrinsic to those qualities). Thus, if the subjective givenness of conscious states is accepted as a phenomenological datum, the subject view turns out to be able to provide in a sense a 'simpler' ontology than the state view (precisely by introducing a second kind of phenomenal properties, which allows not to posit a *sui generis* relation between 'for-ness' and phenomenal qualities). Moreover, once the notion of phenomenal perspective is better clarified, it becomes apparent that it can also play a further significant explanatory role (other than accounting for the existence of subjective character in extrinsic terms).

The notion of phenomenal perspective has been previously characterized as referring to the property in virtue of which an entity can occupy "the standpoint or the position of a [conscious] *person* or *subject*: the 'place' from which they 'see' things" (Crane 2001, 4). A first step in the clarification of this notion may be taken by further developing the analogy with vision suggested in the metaphor above: in the same way in which only a perceiving subject situated at the center of a visual field can be said to have a visual

perspective, a phenomenal perspective can be defined as the property in virtue of which a conscious subject can occupy the center of a phenomenal field; and just as a visual field can be defined as involving a network of visually represented properties, a phenomenal field can be defined as involving a network of phenomenally conscious qualities.

A suggestion along these lines has been put forward by Bayne (2007), who argues that the phenomenal perspectives of subjects should be conceived as explanatorily more fundamental than the phenomenal characters of their mental states precisely because consciousness has a field-based structure. Following Searle (2000), Bayne distinguishes two competing views of how consciousness is constituted – the building block model and the unified field model (2007, 3-4). According to the building block theorist, phenomenal consciousness has an atomistic structure: creature-consciousness is constituted by the phenomenal characters of individual mental states, which can be conceived of as “independent units of consciousness” (Bayne 2007, 4), i.e., as qualitative properties of fine-grained mental states that are conscious independently of what happens to any other co-occurrent state.<sup>38</sup> By contrast, according to the unified field theorist the structure of consciousness is such that the phenomenal characters of individual mental states come into being as aspects of a single phenomenal field and are individuated by abstraction from it.<sup>39</sup> <sup>40</sup> Bayne then suggests that the contrast between the building block model and the unified field model mirrors the contrast between the state view, according to which consciousness should be conceived as being primarily a property of mental states, and the subject view, according to which consciousness should be conceived as being primarily a property of subjects: “the phenomenal field theorist sees [...] creature consciousness as having explanatory priority, whereas the building block theorist accords explanatory priority to [...] (fine-grained) phenomenal states” and thus on the building block model

---

<sup>38</sup> An example of the way in which the building block model can be developed is by defending the idea that any subject is at the same time the subject of “several distinct phenomenal consciousnesses, at least one for each of the senses, running in parallel” (O’Brien & Opie 1998: 387).

<sup>39</sup> As Bayne points out, this conception of the structure of consciousness need not be interpreted as implying that the phenomenal contents of experience do not exist as qualitative properties of distinct fine-grained conscious states, such that “there is only the subject’s total phenomenal field and the various phenomenal features that it subsumes” (2007, 4), but only that those qualitative properties are only phenomenal insofar as they are part of a unified field.

<sup>40</sup> It should be noticed that while the rejection of the atomism of the building block model only entails holism in the general sense that consciousness is something more than the mere numerical sum of phenomenal contents, i.e., it does not entail that the phenomenal contents of experience are necessarily interdependent, i.e., that a certain type of phenomenal quality can only be instantiated if other specific types of qualities are instantiated.

“creature consciousness will be something of an explanatory free-rider on state consciousness” (2007, 6).

The supporter of the state view may object that one’s commitment to the existence of unified phenomenal fields is not sufficient to establish that consciousness should be conceived as being primarily a property of subjects, i.e., that rejecting the atomistic characterization of consciousness proposed by the building block theorist does not necessarily lead to the conclusion that creature-consciousness has explanatory priority over state-consciousness. Granted, phenomenal fields are properties that are ascribed to subjects rather than mental states – hence, in a sense, taking consciousness to be made of unified fields means conceiving consciousness as being a property of subjects. But if that is the only reason why creature-consciousness is taken to be explanatorily prior to state-consciousness, this priority claim could be even shared by building block theorists: fine-grained phenomenal states conceived as independent units of consciousness are ascribed to subjects as well – it is just that they are supposed to make the subject conscious, rather than being made conscious by a property of the subject. That is, independently of whether a fine-grained mental state can become conscious on its own or only as a part of a certain unified field, creature-consciousness will always have some kind of explanatory priority in the sense that devising a full-fledged theory of consciousness requires one to answer the question concerning the nature of the properties making a subject conscious. Ultimately, creature-consciousness is not strictly speaking “an explanatory free-rider on state consciousness” (Bayne 2007, 6) only if the former is ontologically more fundamental than the latter.

While the building block model trivially entails that state-consciousness is explanatorily more fundamental than creature-consciousness, since the building block theorist assumes the ontological priority of the phenomenal properties of mental states, the unified field theorist does not need to take creature-consciousness to be explanatorily more fundamental than state-consciousness, because the adoption of the unified field model does not entail the ontological priority of phenomenal perspectives. That is, even if the basic constituents of conscious experience are taken to be phenomenal fields (instead of individual phenomenal characters existing independently of each other), consciousness may still be conceived as being primarily a property of mental states, for

it may be produced by the phenomenal character of ‘self-unifying’ co-occurrent mental states (constituting consciousness rather than being constituted by it).

For example, the self-representational theory proposed by Kriegel (2009) may be plausibly described as implementing the unified field model while still ascribing explanatory priority to state-consciousness rather than to creature-consciousness. On the one hand, Kriegel explicitly rejects the building block model – as he takes his theory to be “an account of the ontological structure of the global experience/ maximal conscious state”, and claims that “as for sub-maximal conscious states, what makes them conscious is simply that they are logical parts of a maximal conscious state” (2009, 229). But, on the other hand, he also takes the notion of phenomenal character to be explanatorily more fundamental than that of phenomenal perspective, as he holds that a mental state  $M$  is conscious in virtue of having two proper parts,  $M_i$  and  $M_{ii}$ , such that  $M_i$  indirectly represents  $M$  by directly representing  $M_{ii}$ .<sup>41</sup> Since there is no apparent incoherence in the conjunction of these two theses, it seems likely that rejecting the atomism of the building block model and introducing phenomenal fields into our ontology does not compel one to deny that state-consciousness is explanatorily more fundamental than creature-consciousness.

Therefore, taking phenomenal consciousness as being made of unified phenomenal fields does not *ipso facto* mean characterizing phenomenal consciousness as being primarily a property of subjects (though the latter thesis presupposes the former, because of the incompatibility of the building block model with the ascription of ontological and explanatory priority to creature-consciousness). What ultimately determines whether consciousness should be understood primarily in terms of creature-consciousness rather than in terms of state-consciousness is whether or not phenomenal fields are entirely constituted by the network of qualitative properties of mental states that make up the phenomenal contents of experience: creature-consciousness is explanatorily prior to state-consciousness only if the former must be presupposed in one’s explanation of the latter, i.e., if it is the subject’s having a phenomenal perspective that is responsible for the unification of the qualitative properties of mental states into a phenomenal field

---

<sup>41</sup> The only sense in which Kriegel may be said to ascribe explanatory priority to creature-consciousness is that state-consciousness is conceived in terms of the subject’s inner awareness. However, insofar as inner awareness is analysed in terms of the property of mental states of representing themselves, it seems clear that consciousness is conceived as the property of those states of making the subject conscious, rather than as being primarily a property of the subject.

(independently of whether mental states are independently qualitative, as suggested by the modest subject view, or their phenomenal qualitativity is constituted by their being suitably integrated, as suggested by the ambitious subject view).

However, although the ascription of ontological priority to state-consciousness is compatible with the endorsement of the unified field model of consciousness, the combination of the two forces poses a challenging question to the supporter of the state view. Namely, how to conceptualize the relation of phenomenal unity between distinct co-occurrent experiences, resulting in the constitution of a phenomenal field? The supporter of the state view may try to answer by appealing to sub-personal causal processes such as neural synchronicity (e.g., Kriegel 2009, 246), while holding that there is no specific element within phenomenology associated with phenomenal unity – i.e., that these sub-personal processes produce a primitive relation of ‘co-consciousness’ that makes distinct fine-grained conscious states in some way ‘self-unifying’ without being an experience in its own right (e.g., Dainton 2008, 49).

The alternative, for the supporter of the state view, is to appeal to a personal-level process of unification, conceiving phenomenal unity as being a component of the phenomenal character of conscious state<sup>42</sup>, but the viability of this latter strategy is threatened by the so-called ‘just more content’ objection: “how can anything internal to [phenomenal] content determine unity, given that content presupposes unity? What prevents the problem of co-consciousness from applying all over again to it?” (Hurley 1998, 70). By contrast, the supporter of the subject view proposes to conceive phenomenal unity as a structural feature of phenomenology,<sup>43</sup> rather than as an item in it, by grounding its existence in the structure of consciousness, i.e., extrinsic subjective character.

On intrinsic conceptions of subjective character, it is the conscious state’s for-me-ness that is supposed to make the subject conscious of it, hence the supporter of intrinsic higher-order theories cannot explain phenomenal unity in terms of subjective character

---

<sup>42</sup> Kriegel’s own proposal, for example, is that unity is the result of “representations of response-dependent relations among items represented by other items in the phenomenology” (Kriegel 2009, 184).

<sup>43</sup> A suggestion along these lines has been offered by Bayne and Chalmers (2003), who characterize phenomenal unity in terms of the ‘subsumption’ of individual phenomenal states into complex conscious states that have the subsumed mental states as components. The supporter of the subject view, rather than assuming the notion of subsumption as an “intuitive primitive” (Bayne & Chalmers 2003, 40), defines it in terms of the subject’s phenomenal perspective.



(unless he is ready to radically re-define the notion of subject).<sup>44</sup> By contrast, on the subject view, since creature-consciousness is not supposed to be constituted by state-consciousness, it becomes possible to characterize the personal-level process responsible for the existence of phenomenal unity as being a form of subject-involving conscious inner awareness: the subjective character of conscious states is supposed to be constituted by a relation of those states with their subject, hence that very relation can also ground the unity of those states on the unity of oneself as a single subject of experience (i.e., an entity endowed with a phenomenal perspective).

That is, by accounting for the existence of subjective character in extrinsic terms, the subject view offers a ‘top-down’ rather than ‘bottom-up’ personal-level account of phenomenal unity: within the framework of the subject view the formation of unified phenomenal fields will not be characterized as the outcome of the instantiation of a relation (in need of explanation) between distinct co-occurrent experiences, but rather as a result of the same process by means of which consciousness is constituted. Therefore, if consciousness is conceived as being primarily a property of subjects – such that a subject’s phenomenal perspective is not just the property of that subject’s mental states of having a phenomenal character – providing a theory of consciousness will mean *ipso facto* providing a substantial account of the cognitive structure responsible for the unification of heterogeneous phenomenal qualities of distinct mental states into a single phenomenal field.

Taking stock, although intrinsic higher-order theories are specifically devised to acknowledge the hard problem while extrinsic theories are usually associated with eliminativist or illusionist approaches, extrinsic views can also take the hard problem at face value: either by adopting a ‘modest’ view, according to which phenomenal character is made of non-essentially conscious intrinsic qualities that can be made phenomenal by an extrinsic relation of (unconscious) inner awareness, or by adopting an ‘ambitious’ view, according to which phenomenal character is made of non-essentially conscious extrinsic properties of mental states that become intrinsic phenomenal qualities in virtue of the instantiation of an extrinsic relation of (conscious) inner awareness. While the first

---

<sup>44</sup> This is why Kriegel needs to characterize the ontological structure of consciousness as involving *three* fundamental constituents: “(i) a first-order representation, (ii) a higher-order representation of that first-order representation, and (iii) some relationship of cognitive unity between the two, in virtue of which they form a complex” (2009, 233).

option is compatible with the state view, according to which subjects can be said to have a phenomenal perspective simply in virtue of having mental states that possess a phenomenal character, the second option requires the adoption of the subject view, according to which mental states acquire a phenomenal character in virtue of being caught up within the subject's phenomenal perspective.

Moreover, the notion of phenomenal perspective, although generally overlooked or deemed dispensable, is particularly interesting in that it not only allows one to articulate the ambitious extrinsic view, but it may also provide an account of (extrinsic) subjective character and phenomenal unity all at once. The question of how this task could be accomplished will be considered later, as it can only be answered by considering specific versions of higher-order intentionalism.<sup>45</sup> However, before proceeding to analyse specific higher-order theories, it will be argued in the following chapter that the general explanatory strategy presupposed by intrinsic theories is significantly more contentious than the one presupposed by extrinsic theories – as it involves controversial metaphysical assumptions – providing a first reason to those sympathetic with the principles of higher-order intentionalism to prefer extrinsic views over the phenomenal character view.

---

<sup>45</sup> In Part II (§5), it will be suggested that a phenomenal perspective may be constituted in virtue of the subject's instantiation of an 'attention schema' (Graziano 2013), a schematized model of attentional mechanisms, indirectly responsible for embedding first-order states into a subject-object structure which is in turn considered responsible for their acquiring an extrinsic kind of subjective character.

### **3. Metaphysical Implications**

The purpose of this chapter is to argue that the extrinsic view (according to which consciousness is a cognitive mechanism that is extrinsic to conscious states and determines their being conscious) may be preferred over the phenomenal character view (according to which consciousness is made of essentially conscious properties of mental states that make their subject conscious) because of its metaphysical neutrality – contrasting with the phenomenal character view’s commitment to categoricism, a popular but far from uncontroversial metaphysical conception of properties.<sup>46</sup> After a brief presentation of the notions of categorical and dispositional, it will be argued that intrinsic higher-order theories are incompatible with the adoption of a purely dispositional conception of consciousness, insofar as it directly leads to deny that a phenomenal character’s identity is fixed at least in part intrinsically (§3.1). Then, it will be argued that intrinsic higher-order theories cannot escape their commitment to categoricism even by embracing ‘mixed’ views, involving the claim that properties can involve both dispositional and categorical features (§3.2).

#### **3.1. Consciousness as Categorical or Dispositional**

The purpose of this section is to argue that the phenomenal character view entails a conception of consciousness as being, at least in part, categorical in nature. In what follows, after introducing the notions of categorical and dispositional and their role in the contemporary metaphysical debate, it will be presented an argument to that effect – appealing to the incompatibility between the phenomenal character view and the adoption of purely dispositional conceptions of consciousness (§3.1.1). Then, by considering the implications of dispositionalist positions in the debate concerning the metaphysics of fundamental properties (i.e., the properties that ultimately ground reality), it will be argued that this incompatibility follows from the connection between dispositionalism and the thesis that properties’ identities are fixed extrinsically (§3.1.2).

##### ***3.1.1. Categoricalism and dispositionalism***

The distinction between categorical and dispositional properties reflects the difference between the actual, occurrent qualities of objects on the one hand and, on the other hand,

---

<sup>46</sup> Categoricalism, as a general position in metaphysics, has been defended among others by Mill (1967), Ramsey (1978), Armstrong (1997) and Lewis (1999). The list of critics of categoricism includes Shoemaker (1980), Martin (1994; 1997), Bird (1998; 2007), Molnar (2003) and Marmodoro (2009).

the properties that characterize how objects would behave under certain circumstances. For example, while having a certain shape is a manifest property of a glass, characterizing it as it is now, its fragility only concerns how the glass would behave under certain circumstances (as it is the disposition to break when struck). That is, categorical properties can be conceived as non-dispositional since they are essentially occurrent, “here and now, [...] not merely potential features of the objects of which they are qualities” (Heil 2012, 59), whereas dispositional properties “often exist in a ‘dormant’ state here and now as it were, and may (or may not) exercise or manifest their powerfulness when appropriate conditions obtain” (Mayr & Marmodoro 2019). In fact, a dispositional property may exist without ever being exercised and becoming manifest: the glass is fragile independently of whether it eventually breaks, because fragility is essentially the *capacity* to break under certain conditions. In other words, dispositional properties are essentially defined in terms of their intrinsic ‘directedness’ towards their manifestation; thus, they are ontologically independent of the actualization of those manifestations (Molnar 2003, 57). By contrast, categorical properties, being essentially occurrent, are not independent of their manifestations and need not be intrinsically ‘directed’ towards anything – i.e., their identity and individuation conditions may be characterized independently of any causal role they may play.<sup>47</sup>

The metaphysical debate concerning the relationship between categorical and dispositional properties involves a stark contrast between categoricalism, i.e., the view that categorical properties are the fundamental building blocks of reality and thus that dispositional properties (if there are any) are ontologically dependent on their categorical basis (e.g., Armstrong 1997), and dispositionalism, or power monism, i.e., the view that properties are ‘pure powers’ that do not need the help of categorical properties to be anchored to reality, nor to constitute reality itself (e.g., Shoemaker 1979).<sup>48</sup> Within the framework of this debate, the claim that consciousness is a categorical property can be directly derived from the conjunction of the phenomenal character view (P1) with the realist approach to the hard problem naturally suggested by such a conception of consciousness (P2):

P1. Consciousness is constituted by phenomenal character.

---

<sup>47</sup> Whether or not they *should* be characterized as such is a more controversial matter that will be bracketed here.

<sup>48</sup> Intermediate positions will be considered in the next section.

- P2. Phenomenal character cannot have its identity fixed extrinsically.
- P3. Phenomenal character is not a dispositional property. [P2]
- P4. Phenomenal character is a categorical property. [P3]
- ∴ Consciousness is a categorical property. [P1, P4]

The inference from P3 to P4 is justified by the assumption that the non-dispositional nature of a property entails its being categorical (which is true unless properties are conceived as a mixture of categorical and dispositional aspects – a possibility that will be bracketed until the next section), and the inference from the conjunction of P1 and P4 to the conclusion is trivially true: if consciousness is constituted by phenomenal character and phenomenal character is categorical, then consciousness must be categorical as well. Thus, what follows will focus on the inference from the idea that phenomenal character cannot have its identity fixed extrinsically (P2), naturally associated with realism about the hard problem, to the claim that phenomenal character is not dispositional in nature (P3). In particular, it will be argued that if phenomenal character could be characterized in purely dispositional terms, one's theory of consciousness would be compatible with power monism, but that in a world devoid of categorical properties there is no place for phenomenal character intrinsically conceived. For example, if it were the case that, as suggested by supporters of power monism, the bluish quality experienced while consciously seeing the blue sky could be wholly explained in terms of its manifestation conditions and its causal effects, then there would remain no reason to believe that such a quality determines what it is like to consciously see the blue sky in virtue of the way that quality alone, and nothing else, is (i.e., in virtue of its intrinsic features). Thus, it will be argued that unless one is ready to adopt an extrinsic view and embrace eliminativist or illusionist approaches to the hard problem, the extrinsic nature of dispositional properties forces the supporter of the phenomenal character view to assume the categorical nature of consciousness. Accordingly, the inference from P2 (realism the hard problem) to P3 (the non-dispositional nature of phenomenal character) will be justified as follows:

- P2. Phenomenal character cannot have its identity fixed extrinsically.
- P2\*. Dispositional properties have their identity fixed extrinsically.
- P3. Phenomenal character is not a dispositional property. [P2, P2\*]

### 3.1.2. *Dispositionalism and the categorical nature of phenomenal character*

Dispositionalism, or power monism, can be defined as the conjunction of the following two theses:

(1) All fundamental properties are powers [...]

(2) The essence of a power is exhausted by its ability to bring about those manifestations it is capable of producing (Williams 2019, 96).

That is, on dispositionalist views, pure powers are the only properties necessary to constitute our ontology: the list of existent items in the world is entirely constituted by their powerfulness. The case for dispositionalism has some *prima facie* plausibility. For example, even though the would-be power of a glass to break when struck may seem to be grounded on a certain categorical basis, i.e., the molecular structure of the glass, it is not obvious that this apparently categorical structure can be individuated independently of its causal roles and, moreover, it seems that such a structure only obtains as a result of the activity of some more fundamental dispositions (of the molecules themselves, which in turn will depend on the dispositions of their atoms and so on). And if that is the case – if the practice of positing categorical bases for dispositions does not really help explain their ‘powerfulness’ (Marmodoro 2009) – then we may start questioning whether we really need to admit categorical properties into our ontology, as they might just seem superfluous. Moreover, the elimination of categorical properties from the ontology provides a straightforward way of defending the legitimacy of Aristotelian metaphysical categories – such as form, capacity, and essence – that have regained more and more popularity in the last decades (e.g., Fine 1994, Martin & Heil 1999, Molnar 2003, Bird 2007, Marmodoro 2010, Mumford & Anjum 2011) against neo-Humean criticisms.<sup>49</sup> For example, if all properties were powers, then we would have conditions for transworld

---

<sup>49</sup> This is not to say that categoricism is incompatible with neo-Aristotelian views, but only that, differently from dispositionalism, it is compatible with their rejection.

Neo-Humean metaphysics revolves around the idea that distinctness entails freedom from necessary connections (e.g., Mill 1967, Ramsey 1978, Lewis 1999), i.e., that it is not “up for grab whether the ways we individuate objects entail their being separable in nature”, because “there are no metaphysically necessary connections between distinct, intrinsically typed entities” (Williams 2019, 26). Since this great deal of contingency can obtain only if the fundamental building blocks of reality are such that no causal relation they may enter in could contribute to fixing their identity, dispositionalist ontologies are incompatible with neo-Humean metaphysics. That is, if there is no necessity governing the interactions of distinct entities, then their essential (identity-fixing) properties must be non-dispositional in nature – i.e., those entities cannot have metaphysically necessary futures built into them. Thus, those properties must be categorical.

identity for properties (i.e., sameness of dispositional profile), and it would become significantly easier to make sense of the existence of necessary laws of nature and avoid reducing them to meaningless (contingent) patterns, by grounding those laws on the dispositional essence of properties (Bird 2007, 515).<sup>50</sup> However, despite its increasing popularity, power monism is not without critics. In what follows, it will be argued that although the main type of objection against power monism is likely to fall short in demonstrating its incoherence, the implications of the power theorist's response directly provide a justification of the inference from the assumption of a realist approach to the hard problem (P2) to the claim that phenomenal character is not a dispositional property (P3).

Traditional objections to dispositionalism often take the form of supposedly vicious infinite regresses, encountered whenever one tries to determine the individuation conditions and the identity of powers in a world without categorical properties. One such example comes from Lowe (2006, 138; see also Swinburne 1980), who presents the following objection. Given that the identity of a power is determined by the (type of) manifestation toward which it is intrinsically directed, if the manifestation of a power is identical with the instantiation of new powers (as it apparently cannot be anything else, if power monism is true), then the identity of each power is determined by its relations to other powers, and it seems that no property can get its identity fixed. For each property owes its identity to other properties which, in turn, owe their identity to others and so on (see also Bird, 2007, 523-5; Taylor 2018, 1436). Yet, Bird (2007, 526-33) literally shows – through graph-theoretic models – that the problem is only apparent, because the identities of pure powers may be determined in purely relational terms (without appealing to categorical properties) by conceiving them as supervenient on patterns of manifestation relations (see also Williams 2019); and although “there may be structures of powers that are circular (or that involve infinitely many powers) [...] this does not prevent the

---

<sup>50</sup> This may sound as good news to many, since distinctively neo-Humean theses, despite their enduring popularity, have been more and more criticized in the contemporary debate. For example, various philosophers have doubted the legitimacy of inferences going from conceivability to possibility (e.g., van Inwagen 1998, Worley 2003, Howell 2008, Berto and Schoonen 2018) – which allows to reject the thesis that a property's identity is fixed in all possible worlds – and others have put forward significant counterexamples to the counterfactual analysis of causation – presupposed by the neo-Humean characterization of the laws of nature as the result of meaningless activities of pattern-recognition – involving finks and back-up mechanisms (Martin 1994) as well as masking cases (Johnston 1992, Molnar 2003) that may lead to conclude that the right-hand side of the relevant bi-conditionals cannot provide necessary nor sufficient conditions for the possession of the relevant causal powers.

identities of those powers from being fully determined by the asymmetric pattern of those structures” (Bird 2007, 534). Another closely related objection is “the criticism that if everything is just potency, there is not enough actuality in the system” to constitute reality (Bird 2007, 520). That is, rather than focusing on the difficulties in fixing the identity of properties in a world of pure powers, one may try to directly cast doubts on the very possibility that a world of pure powers could exist, i.e., to argue that “dispositions do not have sufficient reality to be genuine properties without the support of something else” (Bird 2007, 521). This objection can be formulated in the form of another (supposedly vicious) infinite regress, since “a power depends for its reality on the manifestation of properties that in turn are powers that depend for their reality on the manifestation of further powers, and those powers depend for their reality upon the manifestation of yet other powers, *etc.*” (Ingthorsson 2012, 531; see also Robinson 1982). But, once again, it is not clear why the mere fact that a power depends for its reality on the manifestation of another power should make the former any less real, once the latter has manifested. Just like the identities of pure powers could be conceived as being supervenient on patterns of manifestation relations, so too their reality may be grounded on those same power structures (since those structures are made of properties that are supposed to be no less real in potentiality than in act). And the fact that power structures must be ontologically more fundamental than the individual powers those structures relate does not seem inconsistent with the fundamental claims of power monism presented above, i.e., that all fundamental properties are powers and that their essence is exhausted by their ability to bring about specific (types of) manifestations.

But although traditional regress-objections seem to fall short in demonstrating the incoherence of power monism, their implications can allow us to justify the inference from the assumption of a realist approach to the hard problem (P2) to the claim that phenomenal character is not a dispositional property (P3). If the identity of any dispositional property must be fixed in terms of its relative position within a certain power structure and the latter is ontologically more fundamental than the former, then it follows that dispositional properties have their identity fixed extrinsically (P2\*). Thus, since the hard problem only arises if we take the phenomenal characters of our mental states to be made of intrinsic qualities, unless one is ready to adopt an extrinsic view and embrace



eliminativism or illusionism, phenomenal character cannot be conceived as a dispositional property.

The supporter of power monism may try to reject P2\* by questioning the traditional characterization of a power's manifestation in terms of the instantiation of further potentialities, and arguing that powers have their identities fixed intrinsically, rather than extrinsically (Marmodoro 2017; 2020), and thus that there is no incompatibility between intrinsic higher-order theories and the conception of consciousness as a dispositional property. The idea that powers have their identity fixed intrinsically stems from the observation that the manifestation of a power can be identified, rather than with the instantiation of other dispositional properties, with its transition from a state of potentiality to a state of activity. And if a power's "manifestation is not the occurrence of a new power" but "simply a different state of the original power: an activated state" (Marmodoro 2017, 59), then it seems that the identity of powers would be partly determined intrinsically. For, although each power's identity would still be fixed by its being directed toward a certain (type of) manifestation, that manifestation would not be conceived as the extrinsic replacement of a potentiality with distinct potentialities, but rather as an intrinsic modification of one and the same power. And if that is the case, P2\* may turn out to be false: perhaps dispositional properties need not be characterized as having their identity fixed extrinsically. In what follows, it will be argued that this objection cannot truly dissolve the incompatibility between dispositionalism and intrinsic conceptions of phenomenal character, insofar the intrinsic qualifications that can be attributed to pure powers offer no help in distinguishing powers from each other – hence, cannot determine a power's identity.

Justification for Marmodoro's alternative characterization of a power's manifestation comes from another traditional objection to dispositionalism, the 'Always Packing, Never Travelling' argument (Martin 1993, Armstrong 1997). Differently from the regress-objections considered earlier, the main concern behind the Always Packing argument is not whether a world of pure powers is logically and metaphysically possible but, rather, whether the kind of reality that pure powers might constitute alone could resemble our own:

Can it be that everything is potency, and act is the mere shifting around of potencies? I would hesitate to say that this involves an actual contradiction. But it is a very counter-intuitive view. [...]

Given a purely dispositional account of properties, particulars would seem to be always re-packing their bags as they change properties, yet never taking a journey from potency to act. For ‘act’, on this view, is no more than a different potency (Armstrong 1997, 80).<sup>51</sup>

The argument can be interpreted as suggesting that standard dispositionalism involves an inaccurate depiction of change: since “the activation of a power in potentiality is merely an instantaneous ‘jump’ to its manifestation, which is another power in potentiality”, it follows that “change is not defined in terms of potency and act, but only in terms of potency to potency” (Marmodoro 2020, 58). That is, a world in which every power’s manifestation is identical with the instantiation of new potentialities would be a world in which no change is ever in act – for, strictly speaking, there would be no such thing as ‘act’ (since a power in act is simply a distinct power). Granted, *some kind* of change could still occur. Even if all properties are powers and every power’s manifestation is just the replacement of a certain potentiality with other potentialities, it seems that this constant journey from potency to potency does constitute real change (though just change of potentialities) – after all, it is not at all implausible to suppose that having a disposition or another can make a difference in the world. But, the objection goes, it would not be the kind of manifest, non-instantaneous change we observe in our reality. Thus, it may seem that, because of the ever-potential nature of dispositional properties, “a world of pure powers can do a great deal, but it falls short of making a world like ours” (Williams 2019, 100). By contrast, if we reject the traditional characterization of a power’s manifestation in terms of the instantiation of further potentialities, by conceiving it as “a different state of the original power: an activated state” (Marmodoro 2017, 59), it becomes possible for the dispositionalist to admit into the ontology something more than pure potencies, i.e., powers in act. And this novel

---

<sup>51</sup> This argument is often interpreted as attempting to put forward a stronger claim, i.e., “the criticism that if everything is just potency, there is not enough actuality in the system” to constitute reality (Bird 2007, 520). But if this interpretation were correct, it would not be clear why Armstrong only regards the idea that “act is the mere shifting around of potencies” as counterintuitive, rather than as incoherent. If powers on their own are not ‘real enough’ because of their ever-potential nature (as suggested by the proponents of the second regress-objection considered above), then it seems that power monism does involve an actual contradiction: the existence of a reality only made of potencies shifting around would be at least logically impossible.

ontological category (that of ‘act’) might be sufficient to account for the manifest, non-instantaneous change we observe in the world while remaining committed to a dispositionalist ontology (Marmodoro 2020).

However, despite the clear formal difference between this conception of powers and their traditional (purely extrinsic) characterization, it is not obvious that the implications of such a difference are strong enough to break the connection between dispositionalism and the controversial thesis that “act is the mere shifting around of potencies” (Armstrong 1997, 80). For, once a power’s manifestation is defined in terms of its transitioning from a state of potency to a state of activity, we can still ask what the difference between those two states consists of. Since within the framework of dispositionalism the difference between a power in potentiality and a power in act cannot be a categorical difference, then it must be a dispositional difference. And while a power in potentiality can have its identity fixed by its intrinsic directedness towards its manifestation (i.e., its state of activity), the identity of a power in act can only be fixed extrinsically, in terms of its directedness towards further potentialities (instantiated in virtue of that power’s manifestation). That is, although on this view a power’s manifestation is not supposed to be identical with those further potentialities, it is still the case that in a world of pure powers we can only determine the identity of that manifestation by appealing to the (types of) further extrinsic potencies it can bring about. Therefore, even though the replacement of a power with other powers will be non-instantaneous (since it will require as an intermediate step the transition of the first power from its potential to its activated state), given that the only difference between a power’s state of potentiality and its state of activity can be a difference in dispositional directedness (towards the state of activity in the first case, and towards the instantiation of further potencies, in the second case), it will still be the case that ‘act’ must be conceived as no more than potencies shifting around. And if that is the case, then this alternative type of dispositionalism may still involve an inaccurate representation of change. That is, the formal difference introduced by rejecting the purely extrinsic characterization of powers, despite its meaningful ontological implications (such as the non-instantaneous conception of change), does not seem significant enough to stop questioning whether the kind of reality that pure powers may constitute alone could resemble our own. If having purely extrinsically defined powers does not get you all the way to our reality because change appears to be something

more than the instantaneous replacement of potency with potency, it is likely that having powers whose identity is partly fixed intrinsically (when in potentiality) and partly fixed extrinsically (when in act) cannot get you all the way to our reality as well, because change also appears to be something more than the *non*-instantaneous replacement of potency with potency:

If all change were like this, then our world would be lacking in character. And our world – as best we can tell – is not characterless. [...] The ‘changes’ [...] are so minimal – so merely formal – that it is impossible to see them as supporting the richness of quality our world appears to embody. It thus looks like we need more than just powers to get some of that character back (Williams 2019, 98; see also Blackburn 1990, Heil 2003).

Clearly, objections of this kind, given their reliance on phenomenological intuitions, can hardly be universally accepted. However, these considerations are certainly relevant for the present purposes – i.e., showing the incompatibility of dispositionalist ontologies with a realist attitude towards the hard problem. If dispositionalism in all its forms entails that the world consists in nothing else than potencies shifting around, then it seems that in a world devoid of categorical properties there would be no place for those intrinsic qualities that are supposed to constitute phenomenal character. That is, the incompatibility between dispositionalism and intrinsic conceptions of phenomenal character was only apparently dissolved by rejecting the purely extrinsic characterization of powers (and thereby denying the truth of P2\*). For, even if dispositional properties are partly characterized in intrinsic terms – by appealing to the directedness of a power in potentiality towards its state of activity – those intrinsic qualifications offer us no help in distinguishing powers from each other and determining their identities. Given that any power in potentiality must be intrinsically directed towards its manifestation (by definition), it is only our (extrinsic) characterization of a power’s manifestation that can help us fix the power’s identity. That is, since we can only determine the identity of a power’s manifestation by appealing to the (types of) further extrinsic potencies it can bring about, and since a power in potentiality is defined in terms of its (intrinsic) directedness towards its extrinsically characterized manifestation, what truly fixes the identity of a power (independently of the state it is in) is only its extrinsic features – for a power’s intrinsic directedness towards its state of activity ultimately amounts to extrinsic directedness, i.e., directedness towards (directedness towards) extrinsic potencies. Thus,

when it comes to determining a power's identity, the intrinsic characterization of that power in potentiality will be just a placeholder for its purely extrinsic characterization (as directedness towards its extrinsically characterized manifestation). And if that is the case, it turns out once again that dispositional properties have their identities fixed extrinsically (P2\*). Accordingly, given that realism about the hard problem presupposes intrinsic characterizations of phenomenal character, it follows that phenomenal character cannot be conceived as a dispositional property unless one is ready to adopt an extrinsic view and embrace eliminativism or illusionism. Thus, if a property can only be either categorical or dispositional, intrinsic higher-order theories require the assumption of categoricism – a metaphysical framework far from being uncontroversial (e.g., Shoemaker 1980; Martin 1994; Bird 2007; Marmodoro 2009). By contrast, insofar as extrinsic views are in principle compatible with dispositionalism, they are *ipso facto* also compatible with categoricism, since even if consciousness is conceived in dispositional terms, it is always possible to hold that those dispositional features are grounded on a certain categorical basis.

The supporter of the phenomenal character view may try to reject this conclusion by questioning the implicit assumption justifying the inference from P3 (that phenomenal character is not a dispositional property) to P4 (that it is a categorical property), i.e., that properties must be either categorical or dispositional. That is, even though the fact that a property escapes purely extrinsic characterizations entails that it is not a 'pure power', it does not obviously warrant the conclusion that it is a categorical property, for it only implies that such a property necessarily has some categorical, non-dispositional features: consciousness may be, at the same time, categorical *and* dispositional. In what follows, it will be argued that although this strategy does in fact allow us to develop viable alternatives to a categorical conception of consciousness (that are also compatible with a realist attitude towards the hard problem), these are not available to the supporter of intrinsic higher-order views.

### 3.2. Consciousness as Categorical and Dispositional

The relationship between the categorical and the dispositional features<sup>52</sup> of a property can be characterized in four different ways:

(a) *Mixed categoricism*:

The categorical features of a property P ground P's dispositional features.

(b) *Mixed dispositionalism*:

The dispositional features of a property P ground P's categorical features.

(c) *Dualism*:

The categorical and dispositional features of a property P are ontologically equally fundamental for the constitution of P.

(d) *Identity view*:

The categorical and dispositional features of a property P are one and the same.

The purpose of this section is to argue that (a), mixed categoricism, is the only viable option for the supporter of the phenomenal character view. First, it will be argued that option (d), the identity view, is incompatible with the phenomenal character view (§3.2.1). Then, it will be argued that option (c), dualism (§3.2.2), and option (b), mixed dispositionalism (§3.2.3), can be properly articulated only with the help of extrinsic conceptions of consciousness – i.e., by rejecting the phenomenal character view. Finally, I will conclude the chapter by suggesting how the adoption of the subject view may help us motivate these viable alternative frameworks (i.e., (b) and (c)).

Before considering the options alternative to (a), it should be noticed that it is not clear how this view could be distinguished from standard categoricism. For the only significant difference between standard and mixed categoricism is that, while in the first framework it is acknowledged that categorical properties may be characterized independently of their causal profiles, in the case of mixed categoricism categorical properties are characterized as necessarily efficacious. And this difference is unlikely to be sufficient to provide a substantial articulation of the thesis that the categorical and the dispositional are not properties in their own right, but only *qua* aspects of 'larger' properties. For, if a property P has some dispositional features D in virtue of its categorical

---

<sup>52</sup> The admittedly problematic word 'feature' is here used to avoid commitment to any particular conception of how the dispositional and the categorical may be related within a single property.

features C (as suggested by mixed categoricism), there is no clear reason why P should not be identified with C (as suggested by standard categoricism), rather than being conceived as C+D. That is, by grounding the dispositional features of a property in its categorical features, one's commitment to the separate existence of those dispositional features become explanatorily irrelevant: the manifestation they are supposed to necessitate (i.e., what they are directed towards) is directly explained by the categorical features of the property. Saying that a property is made of dispositional and categorical aspects supposedly serves to account for its qualitative features in terms of the categorical and for its causal profile in terms of its dispositional features. But once a property's causal profile is grounded on its categorical features, it seems that positing the existence of a distinct dispositional aspect of that property simply becomes superfluous. For example, if a painful mental state (say, the conscious experience of a headache) has its causal profile in virtue of its phenomenal character (e.g., the painful feeling of the headache), it seems likely that, rather than ascribing to that mental state two distinct properties, a quality (e.g., the painful feeling) and a causal profile (e.g., prompting the subject to take an aspirin), one should ascribe the causal profile to the quality itself. If that is correct, then the supporter of the phenomenal character view must explore alternative strategies to hold that consciousness is categorical and dispositional at the same time.

### *3.2.1. Identity views*

The idea that all properties have both categorical and dispositional natures because of an identity relation holding between the categorical and the dispositional has received some support in the last few decades (Martin & Heil 1999, Heil 2003, Strawson 2008). On this view, for example, the allegedly non-dispositional painful quality of one's conscious headache is identical with the mental property playing such-and-such a phenomenal role in one's overall conscious experience (which may be determined, e.g., by the location and the evolution of the headache) and such-and-such causal effects on the subject (such as prompting him to take an aspirin).

Identity theorists generally acknowledge that there is "a seemingly respectable conceptual (if ultimately metaphysically superficial) distinction between an object's categorical and dispositional properties" (Strawson 2008, 274). Yet, they explain away the metaphysical significance of this conceptual distinction by appealing to our capability to grasp and describe a property's essence from different perspectives, or "modes of

consideration” (Heil 2003, p. 119) – in the same way in which, e.g., we can see a duck or a rabbit while looking at a (duck-rabbit) picture that is at the same time a picture of both ((Heil 2003, 119-20; Martin & Heil 1999, 47). Accordingly, identity views involve a threefold claim:<sup>53</sup>

[Identity thesis] For any property P,

P’s dispositionality,  $P_d$ , is P’s qualitativity,  $P_q$ , and each of these is P:

$P_d = P_q = P$  (Heil 2003, 111).

The identity thesis can certainly look appealing to those philosophers wanting to reject the assumption that properties must be either categorical or dispositional. Clearly, one may share the view that the categorical and the dispositional “can be genuinely held apart in thought, but can’t exist apart in concrete reality (Strawson 2008, 271) without assuming the truth of the identity thesis – for the fact that two things cannot exist apart does not obviously entail that they are identical (i.e., they may be distinct though essentially connected). But alternative metaphysical frameworks (i.e., (a), mixed categoricalism, (b), mixed dispositionalism, and (c), dualism), while providing an explanation of *how* the categorical and the dispositional can coexist within a single property, seem to leave an important question open, namely, *why* such a relation holds. By contrast, the identity thesis provides a direct answer to both questions: “identity does the trick, because the two things [i.e., the categorical and the dispositional] are only one thing, and a thing can’t come apart from itself” (Strawson 2008, 272). And if it is true that “the explanation provided by identity is distinguished by the fact that it leaves nothing ‘brute’ or unexplained”, it may seem natural to hold that “the burden of proof lies heavily on those who wish to claim that something other than identity can make it absolutely impossible for two things to come apart” (Strawson 2008, 272). However, this claim can be readily questioned: although identity views may be appealing in that they could seem to possess more explanatory power than their competitors, it does not follow that they are more likely to be true. In other words, the fact that identity views leave nothing ‘brute’ or unexplained is not a ‘truth-indicating’ virtue, such that it lends support to the truth of the identity view, but only a ‘desire-satisfying’ virtue, such that it could make the identity

---

<sup>53</sup> Just like the duck-rabbit picture gives rise to a threefold identity, between the duck-rabbit picture, the duck picture, and the rabbit picture.



view appear preferable to competitors other things being equal.<sup>54</sup> Even granting that the assumption of an identity view is our only hope to answer the question of why there is an essential relation between the categorical and the dispositional,<sup>55</sup> it does not follow that identity views provide an accurate account of how that relationship is articulated. Thus, the burden of proof should be at least equally shared.

The identity theorist, in order to render plausible his account of how the categorical and the dispositional can coexist within a single property, needs to provide positive justification for the claim that we can explain away the metaphysical significance of the conceptual differences between them without simply eliminating from our ontology one or the other. Intuitively plausible identity claims (such as the one concerning the duck-picture and the rabbit-picture) generally do not involve entities that ostensibly possess inconsistent essential features. But if the categorical and the dispositional are one and the same, then a property must be, at the same time, such that its identity may be fixed intrinsically (*qua* categorical) and such that its identity is fixed extrinsically (*qua* dispositional).<sup>56</sup> Thus, the identity theorist owes us at least an explanation as to why two correct descriptions of a certain property – as categorical and as dispositional – can end up characterizing that property in apparently inconsistent ways (given that those two descriptions are *not* supposed to refer to metaphysically distinct aspects of it). And, in order to provide such an explanation, the identity theorist is forced to define the categorical in such a way that, at the ontological level, it is nothing over and above a property's dispositionality (though it may be described differently). That is, if the categorical and the dispositional are to be identified, then properties' identities must be fixed extrinsically:

The identity theory cannot claim that the property's dispositionality is insufficient to make it the thing that it is, because (on this view) the property is itself identical with a dispositionality.

---

<sup>54</sup> The labels for the contrast between these two kinds of theoretical virtue come from Mendelovici (2018, 118).

<sup>55</sup> This claim will be questioned later, by arguing that at least in the case of consciousness alternative metaphysical frameworks may also be able to provide such an answer.

<sup>56</sup> Another apparent inconsistency may be noticed, i.e., that a property must be, at the same time, essentially manifest (*qua* categorical) and ontologically independent of the actualization of the manifestations it is directed towards, i.e., possibly ever-potential (*qua* dispositional). However, dispositional properties can be conceived, in a sense, as essentially manifest – even though their powerfulness may never become manifest. For the fact that a dispositional property is in 'potentiality' does not entail that it cannot be characterized, just like categorical properties, as "here and now", "actual, not merely potential" (Heil 2012, 59), since the term "potential" can be used in its "in its old meaning – 'potent', 'possessing potency or power'", instead of its "second meaning, 'possible as opposed to actual'", and it is clear that "potential properties in the first meaning are of course actual properties" (Strawson 2008, 275).

Given this identity claim, once we have the dispositional in place, we have the whole property: there is no aspect to the property's nature that remains unfixed by its dispositionality because there is an identity here. (Taylor 2018, 1436)

But, if that is the case, the difference between 'pure powers' metaphysics and identity theories seems to vanish. Since on both views any property's identity is fixed extrinsically, it turns out that "there is no notion of 'quality' on which the identity theorist accepts that properties are qualities, but pure powers theorists reject them" because "whatever else is true of qualities, on the identity theory, qualities ultimately are dispositionalities (because they are identical with them)" (Taylor 2018, 1436-7). In fact, categorical qualities are generally defined in very minimal terms by identity theorists:

Ways things *are* are qualities (Heil 2010, 70).

Qualities are categorical [...] here and now, actual, not merely potential, features of the objects of which they are qualities (Heil 2012, 59).

Talking of the distinction [between the dispositional and the qualitative] as being between the dispositional and the categorical can suggest that dispositionality is not really categorical: not really 'there' in the object (Martin 1996, 74).

All being is categorical being because that's what it is to be! (Strawson 2008, 278).

And, within the framework of pure powers metaphysics, dispositional properties are conceived exactly in the same way in which qualities are conceived by identity theorists: powers are characterized as 'actual' and 'here and now' even when in potentiality (fn. 52), they are supposed to be 'really there' in the objects they are ascribed to and to wholly determine the 'ways things are'. This fact may be interpreted as supporting the claim that identity theories collapse onto dispositionalism, or the claim that pure powers metaphysics are really identity theories in disguise – depending on which characterization of the categorical is presupposed. However, deciding which interpretation is correct goes beyond the present purposes. For, on both interpretations, the only way in which the identity thesis can help the supporter of the phenomenal character view in defending the idea that properties are categorical and dispositional is if we choose to define the categorical in such a way that, at the ontological level, it is nothing over and above a property's dispositionality, which in turns leads to accept that any property's identity can

be fixed extrinsically. For example, if it is the case that the categorical painful quality of one's conscious headache is identical with the dispositional property of playing such-and-such a phenomenal role (e.g., a certain intensity and an evolution) and such-and-such causal effects on the subject (such as prompting him to take an aspirin), then the identity of that allegedly categorical painfulness turns out to be fixed in purely extrinsic terms. Hence, identity views are not a viable option for the supporter of the phenomenal character view.

### 3.2.2. *Dualism*

The idea that phenomenal character is constituted by ontologically equally fundamental and distinct categorical and dispositional features can be cashed out in two ways:

*Dualism of types.* Irreducibly categorical properties as well as irreducibly dispositional properties should be admitted into our ontology (e.g., Ellis 2001, Molnar 2003).

In this case, phenomenal character may be conceived as categorical and dispositional by characterizing it as a complex, structured property, resulting from the mereological sum of a dispositional and a categorical property – e.g., the sum of a categorical painful quality, some manifestation conditions (depending on one's preferred theory of consciousness), and a specific phenomenal and causal role.

*Dual-sided view.* All properties are “Janus-like” (Martin 1993, 184), involving irreducibly and ineliminable categorical and dispositional aspects.

In this case, phenomenal character may be conceived as categorical and dispositional by characterizing it as a somewhat ‘simple’ two-sided property, involving distinct categorical and dispositional features that cannot exist independently of each other. For example, a supporter of this view would suggest that the painful quality of a headache and its ability to prompt the subject to take an aspirin are not, in fact, two distinct properties, but rather two distinct features of one and the same mental property.

Both positions have some appealing features as well as some drawbacks. On the one hand, the dual-sided view can be seen as an attempt to put forward a more parsimonious ontology than dualism of types (since it involves only one basic class of properties) while

staying true to its basic intuition (i.e., that we should reject both categoricism as well as dispositionalism). But, on the other hand, the less parsimonious ontology proposed by supporters of dualism of types is better suited than the dual-sided view to provide a positive characterization of the relation holding between the categorical and the dispositional. For, while both positions involve the rejection of dependence relations between the categorical and the dispositional (otherwise, they would collapse onto either categoricism or dispositionalism), dualism of types at least allows us to appeal to mereological composition to describe that relation, whereas the dual-sided view does not, since the categorical and the dispositional are not supposed to be distinct entities which may be combined and disjoined. It seems that the only option for the supporter of the dual-sided view is to characterize the relation between the categorical and the dispositional in terms of supervenience (Giannotti 2019, 612), but this may appear as an unsatisfying strategy insofar as “supervenience itself is not an explanatory relation” in that “it is not a ‘deep’ metaphysical relation; rather, it is a ‘surface’ relation that reports a pattern of property covariation” (Kim 1993, 167) and does not give us any indication as to why that pattern of covariation subsists. However, independently of whether one prefers the non-economical ontology distinctive of dualism of types or the more parsimonious but somewhat mysterious ontology of the dual-sided view, neither can help us in separating the phenomenal character view from the thesis that consciousness is a categorical property – unless one is ready to abandon intrinsic higher-order theories.

The dispositional features of phenomenal character may be characterized as ‘active’ powers, if those features concern what the instantiation of phenomenal character brings about (i.e., the powers *of* phenomenal character), as well as ‘passive’ powers, if those features also concern what it is that brings about the instantiation of phenomenal character (i.e., the power of categorical qualities of becoming conscious, thereby constituting phenomenal characters). That is, a phenomenal character (say, the painful quality of a headache) may be described as having dispositional features for two distinct reasons: because it is disposed to bring about certain effects (such as inducing the subject to take an aspirin), and because all (and only) the categorical qualities it is made of are disposed to become ‘phenomenal’ but cannot constitute phenomenal character alone (i.e., because phenomenal character is constituted by qualities that have a ‘potentially conscious’ nature that they may or may not manifest). In what follows, it will be argued that (i) if we

characterize the dispositional features of phenomenal character as involving only active powers, then either dualism is interpreted as collapsing onto categoricism or as incompatible with intrinsic higher-order theories – unless one is ready to characterize the categorical features of phenomenal character as epiphenomenal; and (ii) if we characterize the dispositional features of phenomenal character as involving also passive powers, then dualism can be regarded as a viable position only within the framework of extrinsic views.

*i.* Phenomenal character as a combination of categorical qualities and active powers.

The relation between the categorical and the dispositional features of phenomenal character may be necessary or contingent. If necessary, epiphenomenalism about categorical qualities is incompatible with dualism but, for the same reason - i.e., because the categorical aspect turns out to be necessitating a certain causal profile – the dualist conception of consciousness may appear as indistinguishable from categorical conceptions:

If the relation [between qualities and powers] is necessary, then it follows that the qualitative side [of a property] necessitates the dispositional side and the latter necessitates the [type of] manifestation [the property is directed towards]. Why not just say that the qualitative side itself necessitates the manifestation? That would make *it* dispositional of course. The power side has become redundant (Molnar 2003, 151).

The supporter of the phenomenal character view may try to avoid this outcome by pointing out that, differently from the case of mixed categoricism, although the categorical aspects of a property can be described as necessitating the manifestation of the property's dispositions, it is also possible to describe the former as being necessitated by the latter. Thus, one may suggest that the relation between the categorical and the dispositional is akin to the relation between the physical and the mental proposed by Russellian Monism – according to which the fundamental constituents of reality are properties of which the physical and the mental are simply aspects. However, although this strategy may be generally available to the supporter of the phenomenal character view, it seems to be incompatible with intrinsic higher-order theories: if the qualitative for-me-ness of conscious states were part of a more fundamental property that also

involves the active powers of phenomenal character necessarily, the power of that property of generating consciousness would not be ascribed to the for-me-ness itself. That is, if the powerfulness of the properties that include phenomenal qualities is not due to the qualities intrinsic features, but rather to the distinct power-aspect of those properties, then consciousness is not conceived as being constituted by the inner awareness produced by a conscious state's for-me-ness but, rather, it would be conceived as being constituted by the power associated with such for-me-ness – thereby defeating the purpose of intrinsic higher-order theories, i.e., taking the intrinsic subjective aspects of a phenomenal quality to constitute consciousness.

If the relation between the categorical and the dispositional features of phenomenal character is contingent – unless one is ready to commit to epiphenomenalism – the supporter of dualism must find a way to characterize the categorical qualities involved in the constitution of phenomenal character as causally relevant (albeit not as having a certain causal profile necessarily). The idea that categorical qualities may be causally relevant without being 'powerful' has received some support in the debate:

Both powers and non-powers are causal difference makers, but not in the same way. The causal difference to an outcome that a power makes depends on, and is explained by, the nature of the power, and the causal difference to an outcome that a non-power makes also depends on, and is explained by, the nature of a power or powers. [...] non-powers are effective but their effectiveness is mediated by the powers there are (Molnar 2003, 165).

However, it is doubtful that this strategy can be applied to the case of phenomenal character. The idea that categorical qualities are effective without being causally operative implies that those qualities can passively interact with powers: just like "powers can 'sense' one another", in that they are "responsive to the presence and the absence of other powers" (Williams 2019, 104), they may also have a built-in "sensitivity" for those categorical qualities. In fact, Molnar argues for the causal relevance of categorical spatiotemporal properties of objects by appealing to the fact that many fundamental physical powers are "location-sensitive", i.e., that "distances between interacting objects, determined by their respective locations, can affect the *outcomes* of the working of powers, without distances or locations themselves being powers" (2003, 164). Similarly, the supporter of dualism about consciousness may try to characterize the dispositional

features of phenomenal character as ‘sensitive’ to the categorical qualities they are (contingently) related to, thereby making the latter partly responsible for what the instantiation of a certain phenomenal character brings about. Yet, as Molnar points out, this (location-) sensitivity must be “written into” the intrinsic directedness of any power: “the sensitivity is inherent, because these differences in manifestations [i.e., in the outcomes of the working of powers] are a consequence of the nature of the power” (2003, 164). And, if that is the case, it follows that those categorical properties that are effective without being causally operative are essential to the determination of a power’s identity (by partly determining the outcome of its manifestation). But since the identity of dispositional properties is fixed extrinsically, it follows that the categorical properties that can be considered as being effective without being causally operative must be properties whose identity is fixed extrinsically as well (such as spatiotemporal properties): powers cannot be sensitive to intrinsic phenomenal qualities without having those intrinsic determinations entering their identity (thereby abandoning dualism for mixed dispositionalism). Therefore, if phenomenal character is conceived as a combination of categorical qualities and active powers, then either phenomenal qualities are conceived as epiphenomenal, or the intrinsic higher-order theorist must conceive phenomenal qualities as ‘powerful’, or causally operative, thereby adopting categoricalism once again – since powerful categorical phenomenal qualities would ground the dispositional features of phenomenal character. The only dualist alternative is to include passive powers into the dispositional features of phenomenal character.

*ii.* Phenomenal character as a combination of powerful qualities and passive powers.

If the dispositional features of phenomenal character include passive powers, it becomes possible to conceive phenomenal qualities as ‘powerful’ or causally operative without giving up the distinctive thesis of dualism about consciousness, i.e., that phenomenal character is constituted by ontologically equally fundamental categorical and dispositional features. For, even though what is brought about by the instantiation of phenomenal character will still be determined by the nature of phenomenal qualities (unless one wants to make them epiphenomenal), if phenomenal character is partly dispositional not only because it is causally efficacious but also because its instantiation depends on the manifestation of the passive power of categorical qualities of becoming

conscious, then having those qualities will be a necessary but not sufficient condition for a mental state to acquire a phenomenal character. Because for every passive power, i.e., “the power to receive change”, there must be an active power, “the power to make change, to which [the former] is responding” (Ellis 2001, 109). Hence, if the dispositional features of phenomenal character include passive powers, it follows that the presence of categorical qualities is just as essential to the constitution of phenomenal character as the presence of the relevant active ‘partner’ power, responsible for the manifestation of the potentially conscious nature of those qualities.

This type of dualism turns out to be incompatible with the phenomenal character view precisely because it requires phenomenal character to be made of categorical qualities that are supposed to be disposed to become conscious (or ‘phenomenal’) but unable to produce phenomenal character alone. If that is the case, then phenomenal character cannot constitute consciousness (as suggested by the supporter of the phenomenal character view), rather, it must be phenomenal character that is in part constituted by consciousness (with the help of potentially conscious categorical qualities). In other words, if the dispositional features of phenomenal character include passive powers, then the constitution of phenomenal character involves, besides the instantiation of categorical qualities that have the power to become conscious, the activity of an extrinsic consciousness-generating power. Therefore, this type of dualist conception of consciousness can only be articulated within the framework of the modest extrinsic view of consciousness, according to which consciousness is the extrinsic property that unveils the qualitative aspects of mental states, thereby turning the intrinsic qualities of mental states into phenomenal characters.<sup>57</sup>

### 3.2.3. *Mixed dispositionalism*

The idea that the dispositional aspects of a property P may ground P’s categorical aspects follows from the idea that although “the essence of a power is exhausted by its ability to bring about those manifestations it is capable of producing” (Williams 2019, 96), the manifestations of (at least some) powers do not only involve the injection of new potentialities into the world but also of new categorical qualities. For example, the manifestation of a consciousness-generating power (say, the exercise of the subject’s

---

<sup>57</sup> The ambitious extrinsic view, according to which consciousness constitutes the qualitative aspects of conscious states, would introduce a priority relation that is inconsistent with dualism.



capacity of inner awareness, directed at the subject's painful headache) may generate, besides newer potentialities (such as being disposed to take an aspirin), also the very same subjectively given qualitiveness of what is experienced as painful. Mixed dispositionalism can be developed in two alternative ways, depending on one's preferred conception of the manifestation of powers: (i) as a replacement of potentiality with other potentialities, or (ii) as a transition from a state of potency to a state of activity.

*i.* Phenomenal characters as qualitative potentialities

According to Williams (2019), we need to invoke qualities to specify the identity conditions of powers (on top of their intrinsic directedness towards their manifestations), in order to account for the qualitative differences we observe whenever some (observable) power is replaced with other (observable) powers. For, as considered earlier (§1.3.2) it seems that if change only involves the (instantaneous) replacement of potency within potency,

There is nothing interestingly different about the states before and after the change. If all change were like this, then our world would be lacking in character. And our world – as best we can tell – is not characterless. [...] The 'changes' [...] are so minimal – so merely formal – that it is impossible to see them as supporting the richness of quality our world appears to embody. It thus looks like we need more than just powers to get some of that character back (Williams 2019, 98).

Hence, Williams concludes that “there must be some sort of difference between the changing states of the world that is not captured by one state's ability to bring about further [dispositional] states (2019, 101). In turn, given the conception of a power's manifestation in terms of the replacement of its potency with further potencies, taking powers to generate qualitative, not purely dispositional changes in the world will entail that those qualities must be conceived as intrinsically determined aspects of powers in potentiality – for qualities ascribed to a power's manifestation will simply be ascribed to the further powers (in potentiality) brought about by that manifestation.

This kind of mixed dispositionalism closely resembles the dual-sided views: although the dispositional is here conceived as ontologically more fundamental than the categorical, on both views properties should be characterized as compounds of dispositional and categorical features. Accordingly, this kind of mixed dispositionalism

inherits some of the difficulties of the dual-sided view in specifying the relation between the categorical and the dispositional. These difficulties are exemplified in the following attempts at providing a description of the nature of that relation:

It is clear enough that a property's being dual-aspect requires a union of the powerful and qualitative within a single property, but it is not obvious what that boils down to. First things first, it is not a matter of dual-aspect properties being conjunctive properties formed out of two other properties. [...] Dual-aspect properties are *composite*, but not composite in the way that tables are composed of molecules, such that they could in principle be removed. They are composite in the way that concrete particulars include properties but are not just collections of properties. It is, to be sure, an abstract form of composition. The two cannot come apart, and the two are not properties in their own right, but they jointly comprise the same property. They are the two aspects of a power property's essence (Williams 2019, 113).

Once again, since the categorical and the dispositional are not supposed to be distinct entities that may be combined and disjoined, the supporter of this type of mixed dispositionalism can only characterize the relation between them in terms of supervenience, giving no indication as to how that that relation subsists. Thus, on this view, it is not clear why the presupposed ontological dependency of the qualitative on the dispositional should be accepted. Moreover, once we have noticed the presence of this close connection between this kind of mixed dispositionalism and dual-sided views, it becomes apparent that the same threats considered earlier will apply as well: mixed dispositionalism can be conceived as a genuine alternative to categoricism (at least, by the supporter of intrinsic higher-order theories) only by committing to an epiphenomenal conception of categorical qualities (Williams 2019, 111-3). For, even though those qualities are conceived as necessary for fully determining a power's identity, characterizing them as causally operative would lead to the same categoricist views considered in the previous sub-section (§1.4.2). If a property's qualitative aspect is taken to be causally efficacious, being necessarily so (since powers have their causal profiles necessarily), one would end up conceiving it as a 'powerful quality' rather than a 'qualitative power', i.e., as a quality grounding the dispositional features of the property it is an aspect of. Therefore, this kind of mixed dispositionalism does not seem to provide new ways for the supporter of the phenomenal character view to reject the thesis that consciousness is a categorical property. That is unless one is ready to reject the

characterization of a power's manifestation as the replacement of potentiality with other potentialities and conceive of it as an intrinsic transition from a state of potency to a state of activity.

*ii.* Phenomenal characters as powers' manifestations

The idea that the "activation of a power" is "an internal 'transition' from one state to another of the very same power", i.e., that a power's "manifestation is not the occurrence of a new power; rather it is simply a different state of the original power: an activated state" (Marmodoro 2017, 59), offers a way of introducing "some sort of difference between the changing states of the world that is not captured by one state's ability to bring about further [dispositional] states (Williams 2019, 101) without thereby ascribing that qualitative difference to powers in potentiality. That is, perhaps qualities show up when dispositions are exercised not because powers enter the world with some (epiphenomenal) qualitative aspects, but because the transition of a power from its state of potentiality to its state of activity involves that power acquiring qualitative features that cannot be manifested until the power is activated. As considered earlier (§1.3.2), once a power's manifestation is defined in terms of its transitioning from a state of potency to a state of activity, we can ask what the difference between those two states consists of. And while the supporter of pure dispositionalism cannot take that difference to be categorical, the supporter of mixed dispositionalism can define the directedness of a power (in potency) towards its manifestation in partly intrinsic terms, rather than in purely extrinsic terms (i.e., only in terms of its directedness towards the further potentialities instantiated in virtue of that power's manifestation), by characterizing a power's manifestation as essentially partly qualitative. Therefore, within the framework of this kind of mixed dispositionalism, phenomenal character may be conceived as categorical and dispositional by characterizing it as the manifestation of a power (consciousness) that, when exercised, constitutes the phenomenal nature of the qualities that make up the contents of conscious experience (or that transforms unconscious qualities into phenomenal ones). And, differently from the previous kind of mixed dispositionalism, this conception of consciousness as categorical and dispositional does not preclude the possibility of considering phenomenal qualities as causally efficacious. For, even though causally operative phenomenal qualities may appear as powerful qualities (i.e., as categorical bases grounding the active powers of phenomenal character), insofar as they

are conceived as part of a power's manifestation, there is no risk of seeing mixed dispositionalism collapse onto a categorical conception of consciousness, since the existence of those powerful qualities will be grounded on the existence of an ontologically more fundamental power (i.e., consciousness).

However, just like the last version of dualism considered earlier, this kind of mixed dispositionalism is incompatible with the phenomenal character view. For, if consciousness were identical with phenomenal character, it would be (wrongly) identified with the manifestation of the consciousness-generating power, rather than with the power itself. If phenomenal character is the result of the manifestation of a consciousness-power, it seems that there is more to consciousness than just the essentially manifest contents of the mental state made conscious (i.e., phenomenal character). Moreover, although this type of mixed dispositionalism is not in principle incompatible with the modest extrinsic view, it seems that its natural implementation is offered by the ambitious extrinsic view: if phenomenal qualitativity of conscious states is constituted by a power's manifestation, there is no clear need to suppose that the properties of mental states are already qualitative before their phenomenal nature is constituted by the manifestation of consciousness.

Taking stock, if the considerations presented in these sections are correct then, unless one is ready to commit to epiphenomenalism, the adoption of the phenomenal character view forces the supporter of intrinsic higher-order theories to assume the controversial thesis that consciousness is a categorical property – since it cannot be a 'pure' power unless one accepts eliminativism or illusionism about the hard problem, and if characterized as being categorical and dispositional, its dispositional features will turn out to be grounded on its categorical features. Alternative metaphysical frameworks compatible with a realist attitude towards the hard problem (i.e., dualism and mixed dispositionalism) can only give rise to higher-order theories with the help of the extrinsic view, i.e., by rejecting the phenomenal character view, while still taking phenomenal qualities to be intrinsic properties.

Before concluding this chapter, however, it should be noticed that the metaphysical implications of the phenomenal character view in no way constitute conclusive reasons to reject that view: as long as one is committed to realism about the hard problem, the idea that consciousness is wholly constituted by essentially conscious intrinsic qualities

certainly appears as a viable possibility. Moreover, it may be even argued that, in fact, since the whole point of the phenomenal character view is that consciousness is constituted by non-structural, purely qualitative properties of mental states, the supporter of the phenomenal character view will gladly welcome the implied commitment to a categorialist conception of properties. Thus, strictly speaking, the essential connection between phenomenal character view and categorialism (and the lack thereof in the case of extrinsic views) should not be even considered as a genuine disadvantage of the phenomenal character view.<sup>58</sup> Yet, even though noticing the metaphysical implications of the phenomenal character view does not lead to the formulation of conclusive objections *against* the view (as someone who finds it plausible with naturally be sympathetic with those implications as well), it may serve at least as a word of caution for any philosopher that finds the higher-order approach to consciousness appealing but does not want to take a side in the historically controversial debate concerning the metaphysics of fundamental properties. That is, if one is interested in trying to understand consciousness in terms of inner awareness and inner awareness in terms of intentionality while suspending one's judgment about the metaphysical nature of what inner awareness makes us conscious of, then one should steer clear of the phenomenal character view and pursue the explanatory strategy proposed by the extrinsic theorist – i.e., the only higher-order strategy that is compatible with the conception of phenomenal qualities as essentially categorial (when such a conception is not accompanied by the further assumption that those qualities constitute consciousness alone), without entailing it.

---

<sup>58</sup> Thanks to Tom McClelland for raising this point.

## **Conclusions to Part I.**

The purpose of this first part of the dissertation was to present the fundamental differences between the conceptions of consciousness involved in the formulation of extrinsic and intrinsic higher-order theories and to argue for the superiority of the explanatory strategy presupposed by extrinsic views. Higher-order theorists share the intuition that consciousness should be conceived in terms of inner awareness of one's own mental life (and that inner awareness is a matter of intentionality) but disagree on whether or not inner awareness is a component of the experienced properties of the mental state that becomes conscious. While the intrinsic higher-order theorist assumes the phenomenal character view, according to which phenomenal consciousness is made of essentially conscious intrinsic properties of mental states that constitute inner awareness and make their subject conscious of them, the extrinsic higher-order theorist rejects the phenomenal character view and claims that phenomenal consciousness is made of non-essentially conscious properties of mental states that are made conscious by their subject's inner awareness. The contrast between these two types of higher-order intentionalism has been considered in terms of their theoretical implications, concerning the various possible approaches to the hard problem, as well as in terms of their metaphysical implications, concerning the possible options available to characterize the relation between categorical and dispositional properties.

Intrinsic higher-order theories, because of their commitment to the thesis that consciousness is intrinsic to the qualities of the mental states we are conscious of, naturally lead to adopt a realist attitude towards the hard problem. In turn, this same commitment also leads to questionable assessments of the approaches to the hard problem alternative to realism. On the one hand, even though the eliminativist about phenomenal character is committed to the existence of consciousness and qualitative properties, his refusal to fix the reference of the explanandum by relying on the subjective point of view of consciousness is naturally interpreted within the framework of the phenomenal character view as leading to target a substantially different phenomenon while giving it the same name (i.e., consciousness). On the other hand, even though the illusionist proposes a seemingly viable position, intermediate between eliminativism and realism about the hard problem, his refusal to rely on the subjective point of view of consciousness for determining the existence conditions of the apparently intrinsic

phenomenal qualities we are conscious of (while using that same point of view to ostensibly fix the reference of the explanandum), is naturally interpreted within the framework of the phenomenal character view as leading to the incoherence of illusionism. By contrast, extrinsic higher-order theories, although usually associated with the rejection of the hard problem, are not only in principle compatible with realist approaches, but may also attempt to solve it (if one assumes an ambitious extrinsic view, according to which inner awareness is conceived as being primarily a property of subjects and entirely constitutes consciousness, instead of a modest extrinsic view, according to which inner awareness does not constitute phenomenal consciousness alone).

Moreover, intrinsic higher-order theories, by conceiving consciousness as intrinsic to phenomenal qualities, are implicitly committed to a categoricist conception of consciousness, and therefore to the rejection of pure dispositionalist ontologies (implying eliminativism or illusionism about the hard problem), as well as of intermediate positions involving the claim that properties can involve both dispositional and categoric features (compatible with realism about the hard problem). By contrast, extrinsic higher-order theories, while being compatible with categoricism, can also allow for alternative metaphysical conceptions of the nature of fundamental properties. And, although this possibility may be deemed irrelevant by the supporter of categoricism, it certainly has significant consequences. For example, by opening the doors to the possibility that phenomenal qualities are grounded on dispositional properties – either because they are categoric qualities endowed with the passive power to be made conscious, or because they are constituted by and part of the manifestation of dispositions – it becomes possible to conceive of consciousness as a capacity of subjects that gives rise to conscious experience when exercised. On the one hand, the supporter of the extrinsic state view may characterize consciousness as the passive power of a subject of acquiring inner awareness (by instantiating an unconscious state that targets the mental state made conscious); on the other hand, the supporter of the subject view may characterize consciousness as the active power of a subject of making conscious the mental states caught up within his phenomenal perspective. In turn, this feature of extrinsic views allows us to explain not only *how* the categoric and the dispositional can coexist within a single property, but also *why* such a relation holds (just like the supporter of identity views): the phenomenal character of a mental state is essentially categoric as well as dispositional because

phenomenal qualities are constituted (at least in part) by the manifestation of a power ascribed to the subject of that state. Furthermore, for similar reasons, extrinsic views are also able to acknowledge the plausible intuition that “when we lose consciousness we do not become a different kind of entity: an unconscious subject is exactly the same kind of entity as a conscious subject” (Dainton 2008, 77), while intrinsic higher-order theories lead to conceive subjectivity as being constituted by the properties of the experience itself, and thus to portray the conscious subject as a fundamentally different entity from the unconscious subject.

Although these theoretical and metaphysical implications of intrinsic higher-order theories may already provide possible reasons why one may prefer the explanatory strategy adopted within the framework of extrinsic theories over the one proposed within the framework of intrinsic theories, none of them can constitute conclusive reasons to reject the intrinsic view – as long as one is committed to realism about the hard problem, the idea that consciousness is made of essentially conscious intrinsic qualities is certainly a viable possibility. However, in the second part of this dissertation, it will be argued that intrinsic higher-order theories, despite being specifically devised in order to take the hard problem seriously, ultimately turn out to be best interpreted as offering a potential solution to the illusion problem instead (i.e., the illusionist’s problem of explaining how consciousness can involve the intrinsic appearance of phenomenal qualities without involving the actual instantiation of intrinsic qualities), thereby defeating their only possible purpose (§4); and that the modest and ambitious extrinsic views presented above offer the most promising higher-order strategies to tackle the hard problem taken at face value (§5).



## **Part II. Varieties of Higher-Order Theories**

### **Introduction**

The second part of this dissertation is devoted to the analysis of the fundamental dimensions of variation among specific higher-order theories and to the assessment of their prospects. In what follows, it will be argued that conceiving of consciousness as a property that is distinct from phenomenal character and at least partly responsible for its constitution, as suggested by extrinsic higher-order views, leads to the development of a variety of promising higher-order theories that cover the whole spectrum of possible approaches to the hard problem, while conceiving of consciousness as a property that is entirely constituted by phenomenal character, as suggested by intrinsic views, leads to the development of higher-order theories that are naturally seen as being devised to take the hard problem at face value but ultimately lead to illusionist positions. The justification for this conclusion will be provided in two steps. First, it will be argued that among representationalist higher-order theories – according to which inner awareness is constituted by the representation of one’s own first-order mental states – intrinsic views offer no significant advantage over extrinsic views in attempting to address the hard problem, and that the case for the former only rests on controversial phenomenological observations (§4). It will then be argued that, while the rejection of the representationalist conception of inner awareness leads the intrinsic theorist to abandon higher-order intentionalism, extrinsic views can provide promising higher-order strategies to tackle the hard problem (§5).

#### 4. Representationalist Higher-Order Theories

Higher-order intentionalism has been defined in Part I as the conjunction of the thesis that the existence of consciousness depends on the subject's inner awareness of her mental states with the thesis that inner awareness is constituted by higher-order intentionality<sup>59</sup>. Representationalist higher-order theories are defined by their commitment to the following articulation of the second thesis: inner awareness is constituted by a higher-order representation of the first-order state made conscious.

The representationalist conception of inner awareness is ordinarily developed within the framework of the state view, according to which consciousness is only derivatively a property of subjects, i.e., there is nothing more to being a conscious subject than having mental states with phenomenal properties. The representationalist state view is compatible with intrinsic as well as extrinsic higher-order theories, and the two are distinguished by their conceptions of the vehicle of the higher-order representation responsible for the constitution of inner awareness. On extrinsic views, since mental states are supposed to be made conscious by inner awareness, the higher-order representation that constitutes consciousness is supposed to be carried by a mental state distinct from the state made conscious. By contrast, on intrinsic views, since inner awareness is conceived as an intrinsic property of the conscious state, the higher-order representation that constitutes consciousness is supposed to be carried by the same mental state that carries the first-order contents made conscious.

The purpose of this chapter is to argue that intrinsic views offer no significant advantage over extrinsic views in attempting to address the hard problem, and that the case for the former only rests on controversial phenomenological observations. The first section (§4.1) will be devoted to the presentation of the fundamental types of representationalist extrinsic theory and the discussion of two major objections concerning their distinctive features. Then, the second section (§4.2) will offer a presentation of the fundamental types of representationalist intrinsic theories and their alleged virtues. Finally, in the third section (§4.3) it will be argued that there are no substantial differences between extrinsic and intrinsic higher-order representationalism with respect to their

---

<sup>59</sup> Intentionality has been defined as the property in virtue of which mental states can exhibit *directedness*, or *aboutness* towards some (intentional) object, property, or state of affairs.

attitude towards the hard problem – as they both naturally lead to illusionism if the intrinsic appearance of phenomenal properties is presupposed.

#### 4.1. Extrinsic Theories

Extrinsic higher-order representationalism can be defined as the conjunction of two theses:

- *Distinctness*. Inner awareness is constituted by a higher-order representational content carried by a mental state distinct from the mental state made conscious.
- *Extrinsicness*. For any first-order mental state, having phenomenal character is an extrinsic property of that state (i.e., being the object of a distinct higher-order representation).

In what follows, after presenting the two possible characterizations of the relevant higher-order states (§4.1.1), it will be argued that objections against extrinsic theories based on criticisms of the distinctness thesis rest on controversial phenomenological observations (§4.1.2), and that even though objections based on criticisms of the extrinsicness thesis do not pose unsurmountable challenges to the extrinsic theorist, they expose the incompatibility of extrinsic higher-order representationalism with a realist approach to the hard problem (§4.1.3).

##### 4.1.1. *HOT vs. HOP*

The supporter of extrinsic higher-order theories may specify the nature of the psychological mode in which the first-order state is represented by appealing to the higher-order thought (HOT) theory (Rosenthal 1986), or to the higher-order perception (HOP) theory, (Armstrong 1980; Lycan 1987). The HOT theory takes inner awareness to be thought-like, while according to the HOP theory inner awareness has a quasi-perceptual nature. Clearly, no HOP theorist believes that we literally have a ‘mind's eye’ which allows us to see our mental states and processes. Inner awareness is characterized as quasi-perceptual insofar as the existence of the relevant higher-order states is supposed to depend on the presence of inner monitoring systems, taken to be functionally equivalent (in important respects) to the systems responsible for external perception. For example, the notion of inner monitoring can be cast in attentional terms: “consciousness is the functioning of internal *attention mechanisms* directed upon lower-order

psychological states and events” (Lycan 2004, 99). Thus, for example, according to the HOP theorist, a subject can consciously see the blue sky only if the (bluish) content of that visual state is a (possibly non-conceptual) represented object of the higher-order mental states realized by these internal monitoring mechanisms. By contrast, within the HOT framework, a mental state’s being conscious is supposed to involve nothing more than the compresence of that first-order state with a higher-order thought about it – such that, e.g., a subject can consciously see the blue sky only if she is thinking about herself as being in that visual state. Thus, it may seem plausible, *prima facie*, that the HOT theory provides a simpler account of consciousness than the HOP theory, insofar as it does not require us to posit the existence of special cognitive mechanisms devised specifically for giving rise to consciousness – for we seem to have independent reasons to suppose that we are able to entertain higher-order thoughts (e.g., Premaek & Woodruff 1978; Wimmer & Perner 1983; Leslie 1987)<sup>60</sup>.

However, not any kind of thought about a mental state can be taken to make that state conscious: some further extra conditions are required. To make a mental state conscious, the relevant higher-order thought must be roughly *simultaneous* with the first-order state, it must be *assertoric*, and it must arise *non-inferentially* (Rosenthal 2005). The simultaneity condition is needed in order to preserve the extensional adequacy of the theory because, e.g., thinking about a pain I experienced last year need not make me feel that pain again. Moreover, the higher-order thought must be assertoric – i.e., it must be ‘belief-like’, affirming that I am in the relevant state – since merely wondering whether one is in a certain mental state should not allow one to be conscious of it. For example, it seems clear that “the strikingly accurate forced-choice guessing that subjects perform in tests for blindsight [...] cannot make the relevant visual states conscious” (Rosenthal 2005, 185). Finally, the non-inferentiality condition is required in order to rule out cases in which a subject discovers herself to be in a certain mental state through being persuaded by means of third-person evidence, such as observation of her behaviour: when arriving

---

<sup>60</sup> An obvious objection to these considerations is that we commonly ascribe consciousness to infants and non-human animals, which are unlikely to possess those same metacognitive abilities considered in the studies referenced here. However, the kind of higher-order thoughts responsible for the existence of consciousness are generally supposed to involve far simpler metacognitive abilities (see also fn. 10). Alternatively, one may bite the bullet and simply conclude that we are mistaken in ascribing consciousness to infants and non-human animals (e.g., Carruthers 2005).

at the realization that, e.g., I am angry because a friend made me notice it, I need not come to feel angry (not on that basis alone, at least, insofar as I might still deny that I am).

The introduction of these extra conditions is particularly significant because not only it allows the HOT theory to avoid obvious counterexamples, but also enables it to capture some of the crucial analogies between perception and inner awareness which, according to HOP theorists, could lead us to the conclusion that the relevant higher-order states are not thought-like but quasi-perceptual. For example, it seems that, just like external perceptual awareness, inner awareness is partly beyond voluntary control – in that we may decide ‘where we look’ but we cannot determine ‘what we see’ when we do (Lycan 2004, 102-6).<sup>61</sup> Thus, the HOP theorist may suggest that we should characterize the higher-order states responsible for the existence of inner awareness as quasi-perceptual (rather than thought-like) precisely because in the same way in which “once the subject has exerted her/his voluntary control in directing sensory attention, e.g., chosen to look in a particular direction or to sniff the air inside a cupboard, the result is up to the world” (Lycan 2004, 105), once inner awareness has been voluntarily directed towards certain aspects of one’s own mental life, the result (i.e., what we are conscious of) is wholly determined by the features of the first-order states we are in (rather than by how we think of them). But the requirement that the relevant higher-order thoughts must be assertoric and simultaneous with the first-order states they are about seems to allow the HOT theorist to provide an equally satisfying explanation as to why inner awareness is partly beyond voluntary control. Just like we cannot deliberately decide what we see when we look at something, we cannot deliberately determine which mental states we believe ourselves to be in: “once the assertoric requirement comes to the fore, our degree of voluntary control seems to shrink if not altogether disappear” and thus “if we construe the voluntariness of higher-order awareness as primarily an attentional matter of where we direct our inner focus, it would seem to favour neither the HOP nor the HOT” theory (Van Gulick 2000, 286-7).

---

<sup>61</sup> This may seem not to be an entirely accurate description of inner awareness, insofar as the “shift of our inner attention often alters, transforms, or even creates the objects that it brings into focus” (Van Gulick 2000, 287). However, these modifications are not *directly* under voluntary control – we can cause them, knowingly, by modulating our attention (which is indeed under voluntary control), but the way in which ‘what we see’ is modified by our attentional focus is still, in a sense, determined independently of our will.

Similarly, the HOP theorist may suggest that the higher-order states responsible for the existence of inner awareness should be conceived as quasi-perceptual because of the subjective immediacy characteristic of inner awareness: just as ordinary perceptual states apparently make us directly aware of external objects and their properties, so too inner awareness seems to make us directly aware of the contents of our own mental states (Van Gulick 2000, 287-8). But, once again, the introduction of the extra-conditions under consideration can allow the HOT theorist to downplay the significance of the analogy between perception and inner awareness: the subjective immediacy of inner awareness can be explained, rather than by positing quasi-perceptual higher-order states, by appealing to the non-inferentiality condition. That is, the fact that the instantiation of the relevant higher-order thoughts does not involve conscious personal-level inferences might explain why inner awareness seems ‘direct’, or not mediated by anything – for the subject is unaware of the process by means of which the first-order state is made conscious (Van Gulick 2000, 290).<sup>62</sup> Therefore, since the extra-conditions under consideration can allow the HOT theorist to capture the analogy between perception and inner awareness without positing the existence of a special inner monitoring system – and considering that, moreover, the analogy is not entirely accurate<sup>63</sup> – it may seem that, once again, the HOT theory can in fact deliver a simpler and more intuitive model of consciousness than the HOP theory.

Yet, significant objections against the supposedly greater simplicity of the HOT theory can be found in the literature. First, the critic of the HOT theory may point out that the introduction of the extra conditions concerning how the relevant higher-order thoughts must be formed (i.e., the simultaneity and the non-inferentiality conditions), needed to capture the analogies between perception and inner awareness and to rule out obvious counterexamples, seems to be logically independent of the higher-order thought analysis of consciousness. That is, the idea that having thoughts about one’s own mental states is what constitutes phenomenal consciousness does not by itself provide any obvious

---

<sup>62</sup> Further observations supporting the analogy between perception and inner awareness have been provided by Lycan (2004, 101-110), but they are similarly unlikely to constitute definitive evidence against the HOT theory.

<sup>63</sup> For example, the variety of sensory modalities characteristic of external perception is absent in the case of inner awareness (Van Gulick 2000, 186). Moreover, while perceptual states apparently involve the instantiation of qualitative properties, the higher-order states responsible for the existence of inner awareness do not (Rosenthal 1997, 740) – though this second observation appears as cogent only under the assumption that perception is not strongly transparent (i.e., that the qualitative properties involved in conscious experience are not only properties of the objects of awareness, but also encompass the specific modality in which those objects are given).

explanation as to why the very same consciousness-generating mechanism should not work for non-simultaneous states and why higher-order thoughts arrived at by means of personal-level inference should not yield the same results produced by non-inferential higher-order thoughts.<sup>64</sup> If what matters for being in a conscious state is having a higher-order thought about it, then it may seem “completely mysterious why a state's having (or lacking) a certain aetiology should be the extra ingredient that turns it into a state that there is something it's like to be in” (Byrne 1997, 123).<sup>65</sup> By contrast, the analogy with perception allows the HOP theorist to hold that his own version of the higher-order theory directly implies that consciousness only arises from representations of simultaneous first-order states whose formation involves no personal-level inferences. For we generally perceive the world ‘as it is now’, in virtue of the fact that perception generally involves causal-informational links between subject and environment, and for the same reason perception seems to be a largely non-inferential process (at least with respect to personal-level, conscious inferences). Thus, the HOP theorist can hold that, similarly, our internal monitors produce higher-order representations of roughly simultaneous mental states and processes without personal-level inferences because they have a quasi-perceptual nature, and perception has those features essentially – thereby preserving the extensional adequacy of the theory (avoiding the counterexamples considered above), without needing to provide further explanations as to why the relevant higher-order representations must have a certain specific path of origin. Therefore, although the account of consciousness provided by the HOT theory may appear simpler than the one provided by the HOP theory – in that it does not require us to posit specific inner-monitoring systems devised to explain consciousness alone – it also seems that the HOP theory may provide a simpler explanation than the HOT theory – in that those conditions concerning the way in which the relevant higher-order representations must be formed directly flow from the core of the HOP theory, while, by contrast, they must be assumed independently of the higher-order thought hypothesis by the HOT theorist (and thus may appear as *ad hoc*).

---

<sup>64</sup> Clearly, these worries do not apply to the condition that the relevant higher-order thoughts must be assertoric, since that condition does not concern how the higher-order thought is formed but rather it specifies an intrinsic feature of the thought itself (i.e., its propositional attitude).

<sup>65</sup> This is not to say that this is a mystery the HOT theorist cannot solve – for example, Rosenthal (2005) appeals to causal connections holding between the first-order and the higher-order states – but only that the higher-order thought analysis of consciousness does not provide such a solution by itself.

Moreover, further doubts about the supposedly greater simplicity of the HOT theory arise when considering the richness of the contents of experience that higher-order thoughts are supposed to make us conscious of. On the one hand, it seems that the content of perception may outrun the representational capacities of thought, i.e., that the concepts that can be deployed in thought may be unable to properly capture every aspect of conscious experience.<sup>66</sup> On the other hand, even granting that the contents of perception do not outrun the representational capacities of thought, it seems that the proposition describing such contents would be, in virtue of its complexity, simply unthinkable. And even if such a proposition were not in principle unthinkable, it seems introspectively implausible that such complex thoughts are ever actually instantiated (Byrne 1997, 117).

The latter two problems – the question of the ‘unthinkable thought’ and the lack of introspective evidence for its existence – may be avoided by appealing to a variety of simpler higher-order thoughts that jointly represent the complex contents of perception (Rosenthal 1997, 743). That is, we do not encounter, in introspection, extremely complex thoughts representing our experiences because we actually instantiate collections of simpler thoughts, each of which represents an aspect of our conscious experience – and those thoughts are simple enough to be individually thinkable. This move, however, raises further problems. First, unless one is ready to give up the phenomenological observation that experience generally appears as synchronically unified, it comes at the cost of diminishing once more the apparent simplicity of the HOT theory’s account of consciousness, insofar as it compels the HOT theorist to provide an explanation of how collections of simpler higher-order thoughts can be connected into a single unified conscious experience (Byrne 1997, 119). Moreover, given the richness of the contents of that unified conscious experience, the quantity of simple higher-order thoughts required to constitute them may simply appear too cognitively demanding (Carruthers 2000, 221-2). The HOT theorist may try to simplify this task by holding that we “need fewer [higher-order thoughts] than might at first appear” because, for example, “the degree of detail we are conscious of in our visual sensations decreases surprisingly rapidly as sensations get

---

<sup>66</sup> It should be noticed that this thesis does not depend on the controversial assumption that perceptual experiences necessarily have non-conceptual content (criticized, for example, by those philosophers defending the cognitive penetrability of perception), but on the significantly less controversial assumption that the concepts that may be involved in perceptual experiences do not exhaust the contents of perception – or at least that the concepts involved in the constitution of perception are more fine-grained than the concepts involved in thoughts about those experiences (McDowell 1994).



farther from the center” and thus it is plausible that “the content of one’s HOTs becomes correspondingly less specific, and that a progressively smaller number of HOTs will refer to successively larger portions of the visual field” (Rosenthal 1997, 743). However, as Rosenthal himself recognizes, it is not clear that this strategy can genuinely help the HOT theorist in explaining how the contents of perception do not outrun the representational capacities of thought. For, even on the (controversial) assumption “that our conscious experience is not as rich in detail as it seems”, in that “phenomenology suggests far more detail than we can actually discriminate”, it is still the case that “the consciousness of our experience is a matter not of what we can discriminate, but of how our qualitative experience seems to us”, and experience “does seem to be richly detailed” (Rosenthal 2004, 25). Accordingly, the best solution for the HOT theorist is likely to involve acknowledging that “no higher-order thought could capture all the subtle variations of sensory quality we consciously experience” and concluding that “higher-order thoughts must refer to sensory states demonstratively, perhaps as occupying this or that position in the relevant sensory field” (Rosenthal 1993, 63).<sup>67</sup> The main problem with this strategy follows from the observation that our ability for demonstrative reference generally depends on perceptual awareness: “in order for me to be capable of thinking, of an item in the world, ‘That object is F’, the object in question must normally be perceptually presented to me” (Carruthers 2000, 223). Thus, similarly, it seems that the possibility of demonstratively referring to one’s own experience and characterizing it as instantiating certain qualitative properties should be grounded on some kind of quasi-perceptual awareness of that experience and its properties (Byrne 1997, 117; Lycan 2004, 101). And, clearly, if that were the case, the HOT theory would simply collapse onto the HOP model, or at least presuppose it. Rosenthal (2005, 188-9; 204-7) tries to overcome this difficulty by appealing to the notion of comparative concepts, claiming that novel experiences can be characterized by means of contrasts with previously encountered ones. For example, according to Rosenthal, demonstratively referring to one’s own visual experience as instantiating a certain shade of red does not require prior awareness of that shade of red

---

<sup>67</sup> This kind of approach could also allow to make justice to the intuition that there are many species of non-human animals able to entertain conscious experiences. Presumably, a dog can have conscious states, such as perceptual experiences and emotions, but it is doubtful that a dog has the conceptual resources for entertaining complex higher-order thoughts (Dretske 1995, 111; Byrne 1997, 112). But if the higher-order thoughts responsible for the existence of consciousness are taken to have a demonstrative nature, it may be possible to argue that the conceptual resources required to have a higher-order thought about a sensory state are meagre enough to suppose that dogs possess them (Rosenthal 1986, 40; 1997, 741-2).

because it can be conceptualized as an experience of a shade of red darker or brighter than others, but it is not clear how comparative concepts could be explanatorily basic, as the formation of a comparative concept seems to always presuppose the possession of some other concept allowing the comparison to be performed (cf. Gennaro 2012, Ch.7).

However, the task of conclusively settling this debate goes beyond the present purposes. For, independently of one's positions on these matters, it seems at least clear that the observation that the HOT theory does not require us to posit the existence of special inner monitoring mechanisms (while nonetheless being able to capture the analogies between perception and inner awareness) is not sufficient to conclude that the HOT theory should be regarded as providing a simpler account of consciousness than the HOP theory. This is not to say that the HOP theory should be preferred only because it does not face the obstacles just considered – as it still faces the important challenge of explaining the nature of the posited inner monitoring system – but only that it is *prima facie* an equally viable strategy to develop the principles of higher-order theories.<sup>68</sup>

#### 4.1.2. *The objection from distinctness*

The thesis that the consciousness-generating higher-order representation is due to a mental state distinct from the state made conscious, shared by the HOP and the HOT theories, is naturally considered problematic by supporters of intrinsic theories, who take subjective character or 'for-me-ness' to be a constitutive feature of phenomenal character (i.e., such that the phenomenal contents of experience do not only involve 'outer' awareness but also awareness of awareness). For, if one's inner awareness of a mental state M is constituted by a distinct higher-order state M\*, then inner awareness cannot be a conscious phenomenon itself (generating 'for-me-ness') without positing the presence of a third higher-order state making M\* conscious, which would presuppose a further level of conscious representation and so on *ad infinitum*. Hence, because of the distinctness thesis, the supporter of extrinsic higher-order theories is forced to characterize the higher-order states responsible for the constitution of inner awareness – hence, inner awareness itself – as ordinarily unconscious. This commitment is often seen as problematic because it requires the higher-order theorist to face the challenge of explaining how we can be conscious of a state's content without being conscious of the

---

<sup>68</sup> These considerations will prove somewhat useful in the next chapter.

state itself: for example, “how can my thinking that I am in pain make me conscious of my pain if I have no idea that I am thinking that I am in pain?” (Rowlands 2001, 301). Similarly, Kriegel (2009, 30) claims that “it is all but incoherent to describe a person as unconsciously conscious of the fact that *p*”, at least in the normal usage of the term ‘conscious’. That is, according to Kriegel, since it is generally accepted that a first-order state’s being conscious involves its subject being conscious of that state’s content, it is not clear how one could dissociate the latter from the former (by conceiving subjects as conscious of the contents of higher-order states without being conscious of the states themselves) without falling into inconsistent use of the word ‘conscious’.

The natural response for the extrinsic theorist is to distinguish two different kinds of awareness, and claim that although there is “*a* way of understanding the concept of awareness such that a person only counts as aware of something if the mental state in virtue of which they are aware of that thing is itself a conscious one”, this is not “the relevant sense of ‘awareness’ which is put to work” by higher-order theories (Carruthers & Gennaro 2020, §7.1).

This line of reply may be seen as favouring the HOP theory over the HOT theory, in that the analogy with perception provides the HOP theorist with an intuitive way of framing the difference between conscious and unconscious awareness: just as unconscious perceptual states ordinarily make us aware (non-consciously) of what they are about,<sup>69</sup> the HOP theorist can hold that unconscious higher-order perceptual states make us aware of the first-order states they represent without being conscious states themselves. By contrast, it may be harder to see how a similar strategy could be adopted within the framework of the HOT theory, for it seems that “with regard to thoughts, if not with regard to perception, the [phenomenal] consciousness of a thought and the transitive consciousness of what it represents go hand in hand: one does not find one without the other” (Rowlands 2001, 304). For example, according to Rowlands, although the

---

<sup>69</sup> Examples include cases of subliminal perception, absent-minded drivers, and dorsal-stream visual representations used to guide action. Clearly, it is possible to deny the plausibility of each of these examples, as the debate concerning the existence and the characterization of unconscious perception is far from being settled. For example, according to Dennett (1991, 137), the case of the absent-minded driver is better conceived as an example of near-instantaneous memory loss rather than as an example of unconscious perception; according to Block (2007), the case of subliminal perception is an example of phenomenally conscious states which are not access-conscious; according to Wu (2020), standard arguments for considering dorsal-stream visual representations as genuine instances of unconscious perception are ultimately unjustified. However, it is reasonable not to let the case against extrinsic representationalist higher-order theories rest only on these controversial objections.

unconscious belief ‘my dog is ill’ may be the cause of various behaviours (such as increased attention toward its needs), it cannot make me aware of my dog’s being ill – “I am doing these things, but have no idea why” – and it is only when the belief that my dog is ill becomes conscious that I can strictly speaking become aware of the dog’s illness: “as soon as I become aware of this content, [...] my thought, of course, becomes a conscious one” (Rowlands 2001, 304). Carruthers and Gennaro (2020, §7.1) suggest that this conclusion may be resisted by challenging the idea that only conscious states can make us aware of what they represent: “Rowlands, when reflecting back on his dog-nurturing behaviour of recent days, could surely conclude something along the lines of, ‘It seems that I have been aware of my dog’s illness all along; that is why I have been behaving as I have’. That is, just as the idea that unconscious perceptions can make one aware of one’s surroundings stems from the observation that those perceptions influence behaviour, the observation that having an unconscious thought with the content ‘my dog is ill’ could make me act as if I knew that my dog was ill may lead one to suppose that such a thought somehow makes me aware of the dog’s illness, though only unconsciously (precisely to account for the dog-nurturing behaviour).<sup>70</sup> Thus, the HOT theorist may appeal to the same analogy the HOP theorist appeals to, and hold that, in the relevant sense of awareness, a subject can be aware of being in a certain mental state in virtue of entertaining a higher-order thought about it while being unaware of the higher-order thought itself.

Yet, even granting the legitimacy of this distinction between types of awareness, the intrinsic theorist can still ask whether unconscious awareness is in fact sufficient to explain consciousness. For, even though one can be made somehow aware of one’s surroundings by having unconscious perceptual states, or of the dog’s illness by having unconscious thoughts about it, a subject is generally not considered conscious of the contents of those mental states only because those contents are (unconsciously) being used to guide action. Thus, the intrinsic theorist may suggest that, analogously, it is not clear why a subject should be considered conscious of a mental state only in virtue of having unconscious inner awareness of that state: if inner awareness only makes us

---

<sup>70</sup> That is, just as one may be aware of the dog being ill without consciously thinking that it is, one may be unconsciously aware of the subjective character of the contents of inner awareness prior to consciously thinking (via either introspection or inference) that one has inner awareness.

unconsciously aware of our mental states, why should one accept as intuitive that it can explain their being phenomenally conscious? That is, once the criteria for awareness are dissociated from consciousness in order to avoid the threat of the infinite regress, the idea that conscious states are mental states one is aware of may appear to lose its intuitive plausibility – as the extrinsic theorist’s exclusion of inner awareness from phenomenology leaves unexplained why one should take the core thesis of higher-order intentionalism to capture a fundamental fact about consciousness:

It is not clear how the presence of an unconscious higher-order state can illuminate the intuitiveness of the idea that every conscious state is a state one is aware of. In general, the presence of unconscious states in us is not available to the folk in a way that makes for intuitiveness. Consider the subpersonal, unconscious visual representations in the dorsal stream of visual cortex, which allegedly control action on the go. Since such states are unconscious, the folk are unaware of their existence, so obviously it is not going to be intuitive that they exist. Even if cognitive science establishes beyond doubt that they do exist, this does not render their existence intuitive. By the same token, since Rosenthal’s higher-order thoughts are unconscious, the folk are unaware of their existence, so it cannot be intuitive that they exist (Zahavi & Kriegel 2016, 51).

According to Zahavi and Kriegel, it is only phenomenological evidence that can explain the intuitiveness of the principle that conscious states are mental states one is aware of, and since the extrinsic theorist takes the higher-order representation constitutive of inner awareness to be unconscious, it follows that he cannot explain why one should regard higher-order intentionalism as intuitively plausible as it is usually depicted by extrinsic higher-order theorists (e.g., Rosenthal 2004, 17).

Yet, the extrinsic theorist may reply that two different issues are being conflated here: the intuitiveness of a principle (i.e., that conscious states are mental states one is aware of) and the intuitiveness of the existence of what makes that principle true (i.e., the presence of unconscious representations of those states). Even though no amount of empirical evidence for the existence of unconscious higher-order representations could explain why it seems intuitive that all conscious states are objects of inner awareness, the principle that consciousness essentially involves inner awareness could be intuitive in virtue of the presence of unconscious higher-order representations that provide indirect

phenomenological evidence for its truth, rather than the direct phenomenological evidence that essentially conscious inner awareness would provide:

Suppose that not all conscious states are consciously represented, but that there are some other phenomenological data, such that the best explanation of those other data is that all conscious states are represented. In these circumstances, those other data would constitute *indirect* phenomenological evidence for the proposition that all conscious states are represented [...] in the sense that what we would have phenomenological evidence for would *not* be the very thing whose existence we were trying to establish (Kriegel 2009, 117-8).

That is, the principle that conscious states are states we are aware of could be justified by phenomenological facts other than the fact that all conscious states are states we are consciously aware of: although those facts may be unable to provide definitive evidence for the truth of that principle, they could at least explain its intuitiveness. A natural strategy for the extrinsic theorist to develop this reply is to hold that although inner awareness is not ordinarily conscious, it can become conscious through introspection – as higher-order representations can be made conscious in the same way in which first-order states are made conscious, i.e., by being the object of a distinct higher-order representation (e.g., Rosenthal 2005, 130). Then, the argument would proceed by induction: deriving the principle that all conscious states are objects of inner awareness from the observation that all introspected conscious states are experienced as objects of inner awareness.

Kriegel (2009, 119) considers this rejoinder but quickly dismisses it – concluding that only direct phenomenological evidence can successfully explain the intuitiveness of the principle that any conscious state is an object of inner awareness.<sup>71</sup> He casts the inductive argument in representational terms: the extrinsic theorist would have to hold that since “all the conscious states one has actively introspected have been represented, it is plausible that all conscious states are represented, including those that have not been actively introspected”; he then argues that “the inductive sample is wildly biased” because “the introspecting itself constitutes the representing” and thus it is trivial that “all the

---

<sup>71</sup> He also considers two other potential sources of evidence, conceptual analysis and philosophical reasoning from first principles, but points out that the former is likely to be fruitful in this context only if some phenomenological data is already presupposed, and the latter can hardly be promising, as “it is not verbally or conceptually true that a state cannot exhibit this property [i.e., consciousness] unless it is represented” (Kriegel 2009, 120) – in fact, Rosenthal (2005) himself also rejects the idea that his higher-order theory should be interpreted as the result of conceptual analysis, and holds that it is better seen as an empirical hypothesis.

conscious states in the sample would be represented if what makes them belong to the sample is that they are introspected” (2009, 119). That is, according to Kriegel, since introspection is characterized by the extrinsic theorist as the representation of one’s conscious states, one has no substantial reason to suppose that the fact that those states are represented depends on something other than their being represented in introspection. Thus, he concludes, the inference from “all introspected conscious states are represented” to “all conscious states are represented” does not resemble ordinary inductive inferences such as the one going from “all observed swans are white” to “all swans are white”, rather, it is analogous to inferring “all swans are observed” from “all observed swans are observed” (Kriegel 2012, 479).

Yet, it is only Kriegel’s reconstruction of the inductive argument (not the argument itself) that leads to this conclusion, as it involves an ambiguous use of the notion of ‘conscious state’ that the extrinsic theorist can promptly disambiguate. The mental states made conscious by being represented in introspection are higher-order representations of first-order states, while the conscious states that are said to be represented in the conclusion are the first-order states themselves: the introspected first-order states are not in the inductive sample just in virtue of being the object of a representation (as suggested by Kriegel), but in virtue of being first-order states represented by higher-order states. Thus, the inductive argument does not resemble the inference from “all observed swans are observed” to “all swans are observed”, because ‘introspected’ and ‘object of a higher-order representation’ cannot be used interchangeably (as the two occurrences of ‘observed’ in the premise of the fallacious inference), since for a first-order state being introspected means being represented as the object of a higher-order representation (i.e., represented as conscious), not just being the object of a higher-order representation (i.e., being conscious). Rather, it resembles the inference from “all observed swans are white” to “all swans are white”, as it proceeds from the phenomenological observation that all introspected first-order conscious states are objects of (conscious) higher-order representations to the conclusion that all first-order conscious states are objects of (unconscious) higher-order representations. Therefore, it seems that the intuitiveness of the principle that conscious states are states one is aware of could in fact be explained by the presence of unconscious higher-order representations providing indirect phenomenological evidence for its truth, insofar as second-order representations can be

made conscious by the unconscious third-order representations responsible for the existence of introspection.

Moreover, even the higher-order theorist who is not willing to endorse Rosenthal's model of introspection<sup>72</sup> could find some other indirect phenomenological evidence for the intuitiveness of the principle that consciousness essentially involves inner awareness, as the presence of unconscious higher-order representations may provide that evidence simply by making one conscious of one's outer awareness (i.e., the contents of first-order states), rather than of the inner awareness itself. This claim has been defended by Coleman (2017), who points out that, in describing conscious inner awareness, the intrinsic theorist may be "misclassifying some more or less subtle feature (or features) of 'first-order' phenomenology" that are directly responsible for the fact that "one can be aware of oneself as a thing that is aware" (2017, 272). In particular, Coleman suggests that inner awareness need not be an item in phenomenology to explain the intuitiveness of the principle that conscious states are states we are conscious of, because one can obtain awareness of having inner awareness as a result of the combination of "the feel of the (pretty routine) conscious thought 'I am aware'" with "the feel of 'self-awareness': the conscious sensory qualities associated with one's own body and mind" (2017, 272). The intrinsic theorist may object that this strategy is self-defeating, insofar as having a conscious thought to the effect that one is aware is simply an expression of one's consciousness of one's own inner awareness. Yet, the possibility of consciously thinking that one is aware does not seem incompatible with the absence of inner awareness from phenomenology:

In being aware of red, I just don't know what my alleged awareness of my awareness of red is meant to feel like; I find only the redness. When you ask me to attend to the relational property of *my being aware of the redness*, still all I find is the redness – I don't seem to enter the picture (in respect of that redness). Of course I *know* I'm aware of redness, since *there it is* for me, subjectively. Similarly, I know there's a camera shooting a television scene, although I can't see the camera, only its output (Coleman 2017, 271).

That is, even without having conscious inner awareness, it is not a big leap to go from the observation that there are conscious qualities present to oneself to the conscious

---

<sup>72</sup> For example, the HOP theorist wanting to avoid positing a second inner monitoring system, or assuming that inner monitors can monitor themselves.



thought ‘I am aware’, as the latter fact may be intuitively inferred from the former. Even though “one feels the qualities, but not that which goes into one’s feeling them”, it seems quite possible that “individuals, in noting the qualities of which they’re aware, can make the trivial (though undoubtedly important) inference that they’re aware” (Coleman 2017, 271). And since this inference can be co-occurrent with phenomenological data, grounding one’s own self-awareness as an individual entity (such as emotions and sensations), it seems possible that having the conscious thought ‘I am aware’ may lead to a further inference, to the effect that “one can be aware of oneself as a thing that is aware” (Coleman 2017, 272) even without having one’s own inner awareness among the items of one’s phenomenology. Therefore, it may be possible for the extrinsic theorist to explain the intuitiveness of the principle that consciousness essentially involves inner awareness even without appealing to Rosenthal’s model of introspection,<sup>73</sup> insofar as he may hold that such a principle is grounded on basic phenomenological observations and “doesn’t require awareness itself to be conscious, any more than *being aware of myself as a thing that is watching a show shot by TV cameras* requires me to see the cameras.” (Coleman 2017, 272).

Taking stock, even though the thesis that the consciousness-generating higher-order representation is carried by a mental state distinct from the state made conscious leads the extrinsic theorist to deny that subjective character or ‘for-me-ness’ is a constitutive aspect of phenomenal character (on pain of regress), it seems that the viability of higher-order intentionalism is not significantly threatened by the extrinsic theorist’s conception of consciousness as unconscious inner awareness of one’s outer awareness – as the core principle of higher-order theories that conscious states are states we are aware of may be justified on phenomenological grounds even without appealing to the direct phenomenological evidence allegedly offered by essentially conscious inner awareness.

The natural rejoinder available to the intrinsic higher-order theorist is to suggest that the appeal to unconscious higher-order representations may be able to explain why we find it intuitive that conscious states are states we are aware of, but cannot explain the reason why that principle is assumed within the framework of higher-order intentionalism

---

<sup>73</sup> Coleman (2017, 272, fn. 87) explicitly rejects it: “I really doubt there’s much to introspection, beyond, perhaps, thinking about the qualities one is aware of. But any sensory (or cognitive) qualities accruing to such thoughts, beyond the qualities of their sensory objects, are not contents the HOT apparatus is charged with generating”.

as offering a sufficient condition for the constitution of consciousness, i.e., why we should suppose that mental states are conscious *in virtue of* our inner awareness. In what follows, this objection will be considered in more detail in relation to the so-called “generality problem” (Kriegel 2009, 143), posed by asking why only mental states can become conscious by being objects of inner awareness, and I will argue that it exposes the incompatibility of the thesis that extrinsic, representational inner awareness constitutes our consciousness of first-order states with a realist approach to the hard problem.

#### 4.1.3. *The generality problem*

Critics of extrinsic theories have raised doubts about the idea that mental states could become conscious in virtue of being objects of unconscious higher-order representations by pointing out that being an object of awareness does not *in general* make conscious the object one is aware of. For example, as Rosenthal himself recognizes, “my being conscious of a stone does not make it conscious” (1997, 738). Thus, it seems that the extrinsic theorist should provide an explanation as to why only mental states can be made conscious by becoming objects of awareness: since “a rock does not become conscious when someone has a [thought<sup>74</sup>] about it”, one may ask “why should a first-order psychological state become conscious simply by having a [thought] about it?” (Goldman 1993, 366). And although the extrinsic theorist may reply that it is an analytic truth that only mental states can be conscious (Byrne 1997, 111), i.e., that the generality problem does not apply because the word ‘conscious’ is only applied with respect to mental states (Lycan 1996, 758-9), it seems that such a move would simply leave the generality problem unanswered:

If my perceiving or thinking of a stone, a pencil or my nose does not turn any of them into conscious objects, then why should my thinking of one of my own mental states suffice to make it conscious? And even if it suffices as a matter of extensional adequacy, *what is it about the relevant intentional relation or its correlates that explains why it does so in the meta-mental case* even though perceiving or thinking of some object *x* does not in general make *x* conscious. It will not do to simply appeal to usage and say, “We apply ‘conscious’ to mental states we know of but not to others things we know.” What is required is some explanation of why we do so, some

---

<sup>74</sup> In the original passage, Goldman uses the word ‘belief’ instead, but such a formulation of the generality problem would presuppose an incorrect characterization of the target theory.

account of the intuitive difference we feel between the two cases which grounds the difference in usage (Van Gulick 2004, 72; italics mine).

As suggested by Van Gulick, providing an answer to the generality problem means explaining what renders inner awareness of mental states essentially different from any other instance of awareness, or what renders mental states essentially different from any other object of awareness. Extrinsic theorists may try to explain both at once. On the one hand, they can reject the general claim that being an object of awareness is sufficient for something's being conscious because, according to higher-order theories, things like stones are simply not the kind of entity that we can be aware of *being in*. On the other hand, they can argue that, since conscious states are conceived as states we can be aware of being in, representational relations with one's mental states are essentially different from representational relations with external objects in virtue of the implicit subjective aspect of the former – that may become conscious in virtue of introspection or as the content of a conscious thought.

But the intrinsic theorist may still ask: why should one suppose that mental states are the only kind of entity we can be aware of being in? After all, one can have higher-order thoughts with the content <I am in *this* state>, independently of whether the relevant state is mental or non-mental. And since according to the extrinsic theorist a mental state's being conscious amounts to that state being the object of this kind of higher-order representation, it seems that he should also allow that non-mental states can become conscious when made objects of analogous higher-order representations. Yet, it is evident that “even if I were to come to know about states of my liver noninferentially and nonobservationally [...], that wouldn't make those states P-conscious [i.e., phenomenally conscious]” (Block 1995, 280). Thus, the intrinsic theorist may argue, it seems that the extrinsic theorist cannot simply address the generality problem by relying on the features of inner awareness alone (i.e., its being a higher-order representation of a state the subject is in), but rather he needs to appeal to features of mental states that make the awareness of those states essentially different from any other instance of awareness (Dretske 1995, 97). Still, given the thesis that a mental state's being conscious is constituted by the higher-order representation of that state, it is doubtful that a mental state's having some distinctive feature could be relevant to the constitution of consciousness:

If [the higher-order state]  $M^*$  gave rise to consciousness by modifying [the first-order state]  $M$ , then it would make a difference what characteristics  $M$  has (for example, being mental). It could be claimed that only states with the right characteristics can be modified by being appropriately represented in such a way as to become conscious. But then we must keep in mind that, according to higher-order theory, conscious states do *not* undergo any (non-Cambridge) change in response to the fact that they are appropriately represented. Their being conscious is *constituted* by their being appropriately represented. It is not so clear, then, what difference it makes whether an internal state has certain characteristics or not (Kriegel 2009, 143-4).

However, even though no feature of the state made conscious can play an essential role in the constitution of consciousness, the extrinsic theorist may still resist the conclusion that non-mental internal states are potentially conscious in the same way in which unconscious mental states are. For higher-order representations of mental and non-mental states may differ in virtue of specific features of the higher-order representations themselves rather than in virtue of specific features of the referents of those representations (i.e., the states made conscious). Just as representational relations with external objects differ from representational relations with one's mental states because the former lack the (potentially conscious) implicit subjective character of the latter (as they are not higher-order representations), representational relations with non-mental internal states may lack the right kind of subjective character that is essential to consciousness. Namely, the higher-order representation of a non-mental internal state can only be subjective in the sense that it can make one aware of that state *being in oneself*, whereas the higher-order representation of a mental state is also subjective in the sense that it can make one aware of that state being *for oneself* (i.e., its being subjectively given). Compare: when an unconscious higher-order representation with the content <I am in *this* state> targets the state of one's liver, it could only make the subject unconsciously aware of having a liver that is in a certain state (i.e., the represented state is only ascribed to a part of the subject); by contrast, when an unconscious higher-order representation with the content <I am in *this* state> targets one's mental states, it can make the subject unconsciously aware of being the entity to whom the state is given (i.e., the representational state is directly ascribed to the subject, whereas the state of one's liver can only be ascribed to oneself in virtue of the liver's being physically contained within one's body). Thus, it seems that the extrinsic theorist may address the generality problem

even without appealing to specific features of the states made conscious, by focusing on the differences between representing something as a property of a part of oneself (such as the state of one's liver) and representing something as a part of one's mental life (i.e., as a property of oneself as an individual).

Yet, addressing the generality problem as suggested above puts significant constraints on the attitude towards the hard problem available to the extrinsic higher-order representationalist. For assuming a realist attitude towards the hard problem means accepting that the difference between mental and non-mental entities is at least partly explained by the fact that the mental states made conscious have intrinsic phenomenal properties that non-mental entities cannot have. But, on the one hand, the extrinsic theorist cannot take the intrinsic features of phenomenal properties to consist in their subjective character (on pain of regress) and, on the other hand, the extrinsic theorist also cannot take the hard problem at face value by accepting that the mental state made conscious has an intrinsic qualitative character, as no feature of the state made conscious is supposed to play an essential role in the constitution of consciousness (since a first-order state's being conscious allegedly amounts to its being the object of a higher-order representation). Thus, even though "it doesn't follow immediately from the fact, assuming it to be a fact, that *reddishness* is an intrinsic property, that consciousness is also an intrinsic property" (Levine 2001, 107), within the framework of higher-order representationalism phenomenal character could not be intrinsic unless consciousness were intrinsic as well. That is, given that the first-order state's intrinsic qualities (if there are any) cannot play a constitutive role in giving rise to phenomenal character (because of the extrinsicness assumption), and given that the mental state made conscious cannot have intrinsic form-ness (because of the distinctness assumption), it follows that the extrinsic higher-order representationalist cannot ascribe intrinsic phenomenal properties to the conscious state. Therefore, the extrinsic theorist can either embrace eliminativism about the hard problem (e.g., Rosenthal 2005), or assume illusionism, acknowledging that the mental states made conscious appear as having intrinsic qualities but conclude that their being conscious is a purely extrinsic matter (e.g., Graziano 2013).

Taking stock, although the generality problem does not provide conclusive evidence against extrinsic higher-order representationalism, it can point at an (apparently) attractive feature of intrinsic theories: if one wants to explain inner awareness in terms of

higher-order representation and at the same time acknowledge the reality of the hard problem, rejecting the thesis that inner awareness is extrinsic to the conscious state seems to be the only viable option. The purpose of the remainder of this chapter is to argue that the promise of intrinsic theories to take the hard problem seriously is ultimately unfounded, because the intrinsic representationalist conception of inner awareness naturally leads to the same illusionist characterization of consciousness proposed by their extrinsic counterparts (who accept that conscious states involve the appearance of intrinsic phenomenal properties). Before defending this claim, however, intrinsic theories will be presented in more detail.

## 4.2. Intrinsic Theories

According to intrinsic higher-order representationalism, inner awareness is not constituted by an unconscious higher-order mental state  $M^*$  distinct from the first-order state  $M$  made conscious; rather, the relevant higher-order representation is characterized as an integral part of the conscious state itself – which is in turn conceived as essentially involving first-order (world-representing) aspects as well as higher-order aspects (representing the experience itself). The intrinsic conception of inner awareness may be in principle developed in two alternative ways, depending on one's preferred characterization of the relation between the two components of conscious states,  $M$  and  $M^*$ :

*Identity relation.* A mental state  $M$  is conscious if and only if  $M$  is identical with  $M^*$ : a first-order state becomes conscious in virtue of acquiring a higher-order content that presents its first-order content to the subject by representing it.

*Mereological relation.* A mental state  $M$  is conscious if and only if it is part of a complex state, constituted by the sum of  $M$  and  $M^*$ : a first-order state becomes conscious in virtue of being integrated with a higher-order representation that presents it to the subject by representing it.

The intrinsic theorist wanting to describe the relation between the mental state made conscious and its higher-order representation in terms of identity cannot explain that relation in purely causal terms (i.e., arguing that a mental state can cause itself to self-represent), since causation is generally understood as an anti-reflexive relation while self-representation is essentially reflexive: since “no event, fact, or state can cause its own

occurrence”, it follows that “the simple causal–covariational account renders self-representation impossible (Kriegel 2009, 206). Thus, the only option for the intrinsic theorist who takes M and M\* to be identical<sup>75</sup> is to explain self-representation in terms of a specific functional role possessed by mental states who come to represent themselves. An articulation of such a strategy has been put forward by Carruthers (2000; 2005; 2019), according to whom the existence of conscious experience depends on the availability of first-order states to an extrinsic HOT-producing faculty (i.e., the mind-reading system). Carruthers initially named his theory as ‘Dispositionalist HOT theory’, since no occurrent HOT must be presently targeting the mental state made conscious in order to give rise to consciousness – it is only required that the relevant first-order state is held in a “short-term memory store” which can be accessed by “down-stream concept-wielding consumer systems” (2000, 241). However, he later renamed it ‘Dual-Content theory’ (Carruthers 2005), better capturing his commitment to the intrinsic view: inner awareness is in fact conceived as constituted by the higher-order content acquired by first-order states in virtue of their availability to the mind reading system, i.e., although such availability is essential to the constitution of consciousness, it is not identified with consciousness itself – which is supposed to consist instead in the addition of experience-representing contents to world-representing mental states: “the very same perceptual states which represent the world to us [...] can at the same time represent the fact that those aspects of the world [...] are being perceived” (Carruthers 2000, 242).

The natural objection to a theory of this kind is that just as the mere disposition to see the sky does not cause one to actually see it, the mere disposition to think about one’s mental states should not give rise to consciousness of those states: “how can something which hasn’t actually happened to a perceptual state (namely, being targeted by a HOT) confer on it – categorically – the dimension of subjectivity?” (Carruthers 2000, 235).<sup>76</sup> Carruthers addresses this objection by appealing to consumer semantics and claiming that a first-order state’s availability to the mind-reading system can enrich the very content of that state: “consumer semantics only requires dispositions to make judgements or

---

<sup>75</sup> Unless one is ready to abandon the representationalist’s hope for naturalization, which would render somewhat pointless the appeal to the notion of representation in the first place.

<sup>76</sup> Clearly, there are non-dispositional facts that, on this view, matter for the constitution of consciousness – namely, the presence of first-order states in short-term memory, which may be understood in terms of their occurrent ‘broadcasting’ in a global workspace (Carruthers 2019). However, it is still not clear how higher-order intentionality can enter the picture, except in dispositional terms.

inferences, in order for the content of the consumed state to be determined”, hence, there is no reason “for insisting that the HOTs which are involved in the explanation of phenomenal consciousness must be actual ones” (Carruthers 2000, 241). However, even though there is a clear sense in which “anything which I may be disposed to believe, immediately and non-inferentially [...] may be said to be something which I know (dispositionally), or of which I am aware” (Carruthers 2000, 234), one may still doubt whether it is in this sense that we ‘know’ that we are aware in consciousness. Ultimately, the only sense in which conscious states have a phenomenal character on Carruthers’ theory is that those states “possess properties of a sort which can be available for immediate introspective recognition”, i.e., “the feel of a mental state is that property in virtue of which we can recognise it when we have it” (2000, 255), but is unlikely that such a feel can essentially involve conscious subjective character when introspective capacities are not exercised. Thus, it seems clear that the intrinsic theorist cannot choose this strategy to maintain his promise of taking the hard problem at face value: conceiving of consciousness as being a part of the conscious state does not *ipso facto* mean that phenomenal properties can be conceived as intrinsic properties of the mental state made conscious. Accordingly, the intrinsic theorist wanting to take advantage of the limitations of extrinsic theories brought to light while considering the generality problem should prefer a mereological account of the relationship between the first-order and the higher-order states involved in the constitution of consciousness.

An eminent example of an intrinsic mereological theory has been offered by Kriegel’s (2009) ‘Self-Representationalism’.<sup>77</sup> On this view, the existence of conscious experience depends on the integration of a first-order state into a complex mental state involving its higher-order representation – conceived as a conscious feature of experience (i.e., the subjective character or for-me-ness of conscious qualities), such that (e.g.,) the conscious

---

<sup>77</sup> Gennaro’s (1996, 2012) ‘Wide Intrinsicity View’ is also presented – as suggested by the name of the theory – as another possible mereological intrinsic view: conscious states are “are individuated widely so as to treat the meta-psychological state as intrinsic to the conscious mental state” (1996, p. 16). However, on this account, consciousness is not taken to be intrinsic in the sense under consideration, i.e., as intrinsic to the phenomenal character of the mental state made conscious. Rather, according to Gennaro, the constitution of consciousness involves the presence of unconscious higher-order thoughts targeting first-order states as suggested by extrinsic theorists but, rather than identifying the mental state made conscious with the first-order state alone, the resulting conscious state is considered as a complex entity that involves the higher-order state as well. The obvious objection to this kind of proposal is that it makes the distinction between intrinsic and extrinsic views somewhat arbitrary, since it provides no substantial reason to conceive the first-order state made conscious and its higher-order representation as parts of a single mental state (Kriegel 2009, 223-4; Coleman 2015, 2720-1).



visual state representing the blue sky involves two elements, “the perception of blue and the awareness of that perception”, which “are unified [in a single mental state] by some psychologically real relation whose dissolution would entail the destruction of the experience” (Kriegel 2009, 222). This psychologically real relation – allegedly experienced as the synchronic unity of conscious experience (2009, 233) – is ultimately characterized by Kriegel in sub-personal terms, by appealing to cognitive processes of information integration obtained through neural synchronicity (2009, 246). Thus, once the unity of the represented and the representer is justified by claiming that the subjective character of conscious states is a fundamental aspect of their phenomenal character, it becomes possible for the intrinsic theorist to argue that the first-order state made conscious and its higher-order representation are part of one and the same conscious state because they form a mereological *complex*, whose parts are essentially interconnected, rather than a simple mereological *sum* whose parts are only contingently tied to each other and could give rise to a conscious experience even if their relation were broken (Kriegel 2009, 221).<sup>78</sup>

The integration of the higher-order representation with the first-order state made conscious allows the higher-order theorist to recognize the existence of subjective character as a constitutive feature of phenomenal character without falling prey to the infinite regress of conscious states that threatens extrinsic theories: no further level of representation needs to be introduced, because higher-order representations are supposed to become conscious *qua* part of conscious states. And even though Self-Representationalism must still face a similar kind of infinite regress, concerning conscious representational properties instead of conscious states, it can be quite easily addressed. Clearly, in order to conceive the relevant higher-order representation as genuinely conscious (rather than as an unconscious aspect of a conscious state as suggested by Gennaro) without directly ascribing its content to the first-order state made conscious (as suggested by Carruthers), the conscious state must represent *all* of its

---

<sup>78</sup> That is, while Gennaro’s Wide Intrinsicity View offers no reason to believe that, were the connection between the parts of the widely individuated conscious state broken, the conscious state would simply go out of existence – as the first-order state would cease to be conscious only if the higher-order state disappeared – Kriegel’s Self-Representationalism, by appealing to conscious inner awareness, offers a reason to suppose that breaking the connection between first-order and higher-order state is enough to destroy the whole conscious state, as the (allegedly) experienced subjective character associated with the first-order state’s phenomenal qualities could not present those qualities to the subject anymore. The question of whether Kriegel’s conscious states really are complex rather than sums will be considered in the following section.

representational properties, including its higher-order ones. Yet, if the conscious state does not only represent its first-order parts but also its higher-order parts, it seems that the conscious state consciously represents itself as representing itself. Thus, once again, it may seem that there needs to be a further level of self-representation in virtue of which it is so, and so on *ad infinitum* (Nida-Rümelin 2014, 278). However, the self-representationalist can stop the regress precisely by pointing out that the higher-order representation is an integral part of the conscious state, and thus it could *indirectly* represent the whole conscious state by directly representing only its first-order part – just like a painting can indirectly represent an entire house by directly representing only its front. And despite the inherent difficulties in specifying exactly the conditions of possibility of indirect representation,<sup>79</sup> it seems likely that if conscious states are in fact complex states whose first-order and higher-order parts are jointly necessary and sufficient for the constitution of conscious experience, the first-order state amounts to such a significant portion of the whole that the latter can be indirectly represented by means of a representation of the former (Kriegel 2009, 225-7).

In turn, the possibility of conceiving the relevant higher-order representation as a conscious aspect of the conscious state directly provides the intrinsic theorist with a solution to the generality problem that need not involve any appeal to intrinsic qualitative character, nor the commitment to eliminativist or illusionist positions: conscious states have an intrinsic subjective character that non-mental entities cannot acquire. That is, being the object of an intentional mental state can only make mental states conscious because, on intrinsic higher-order theories, only mental states can instantiate the internal representational relation required to give rise to conscious experiences (i.e., the relation between first-order contents and higher-order representations of them, together with the indirect representation of the latter). Thus, it seemingly becomes possible for the higher-order representationalist to take the hard problem at face value without renouncing the sufficiency of the higher-order analysis of consciousness. It will be argued in what follows, however, that this is a mistaken impression.

---

<sup>79</sup> The fact that something is part of a larger whole does not obviously imply that a representation of the former counts as an indirect representation of the latter – a table is part of the world, but no representation of a table could be conceived as an indirect representation of the whole world. Plausibly, in order for some representation to count as an indirect representation of something else it is required that the represented part amounts to a significant portion of the whole, and that it is well integrated into it. Both conditions, however, may be subject to some degree of vagueness (Kriegel 2009, 225-7).

### 4.3. The Implications of Representationalism

The purpose of this section is to argue that intrinsic higher-order representationalist theories, except for their different description of phenomenology (as essentially involving subjective character), are more akin to their extrinsic counterparts than they are often portrayed to be. First, it will be argued that both intrinsic and extrinsic theories are committed to the claim that the first-order states made conscious by inner awareness may not exist, and thus are not essentially involved in the constitution of consciousness (§4.3.1). Then, the chapter will be concluded by claiming that, for this very reason, accepting the intrinsic appearance of the phenomenal properties of the states made conscious naturally leads both intrinsic and extrinsic theories to conceive of consciousness as an illusion – thereby taking away the supposed advantage of intrinsic theories, i.e., their ability to take the hard problem seriously (§4.3.2).

#### 4.3.1. *The objection from intimacy and the constituting representation view*

Intrinsic and extrinsic varieties of higher-order representationalism need to explain why, when considering our inner awareness of conscious states, “there seems to be a more intimate cognitive connection between the subject and what she is conscious of, or the consciousness itself than is present in other circumstances” (Levine 2006, 177). That is, it seems intuitively plausible that the connection between subject and experience established by inner awareness appears to be more intimate than the one relating a representing subject and a represented object, in that “the properties of experience are [...] cognitively present to us” and “we stand in a kind of epistemic relation to them that is more intimate, more substantive, than the kind of relation that obtains between our minds and other items” (Levine 2006, 159).

One intuitive way to spell out the objection from intimacy is to point out that it seems subjectively “that there is no gap to begin with between the awareness and what one is aware of” (Kriegel 2009, 109), i.e., that consciousness appears to provide us with direct, immediate access to the mental states made conscious. If the intimacy datum is interpreted along these lines, the intrinsic theorist has a straightforward reply: since the conscious state includes the first-order state as well as its higher-order representation and consciousness of the former and consciousness of the latter are essentially tied to each

other in virtue of indirect self-representation, it seems that “there is no causal process that mediates the formation of S’s awareness of M [i.e., the resulting conscious state]: M comes with the awareness of it” (Kriegel 2009, 154). However, although this kind of answer is not available to the extrinsic theorist, the objection from intimacy can be addressed without any need to appeal to conscious inner awareness. For the claim that the relation between our inner awareness and what it makes us aware of is more intimate than any representational relation depends on our subjective experience of it. Thus, the extrinsic theorist may point out that accounting for the intimacy characteristic of inner awareness does not require us to explain why there is no gap between our inner awareness and what it gets us acquainted with, but only why it seems so: even if “we’re conscious of our conscious states in a way that seems, subjectively, to be direct and unmediated”, it is possible that “something might mediate even though we are subjectively unaware of anything doing so”, and therefore “the datum we need to explain is not actual immediacy, but rather subjective immediacy” (Rosenthal 2004, 33). And, in this sense, the extrinsic theorist may well be able to provide a reasonable explanation of the intimacy datum. Precisely because the higher-order representations responsible for our inner awareness are supposed to be unconscious, we are not aware of the higher-order representations themselves, but only of what they are about; and, moreover, since the formation of the relevant higher-order representations does not involve personal-level inferences, the subject is unaware of the whole process by means of which the first-order state is made conscious.

Yet, the intimacy of the relationship between our inner awareness and what it makes us aware of may also be characterized in epistemic terms (rather than only in terms of immediacy): our acquaintance with the phenomenal character of our conscious states seems to have an epistemically privileged status, such that it does not admit the possibility of an appearance-reality distinction. That is, it seems plausible that the identity of the phenomenal character of an experience, i.e., its qualitative character, is necessarily fixed by the way in which that experience subjectively appears to be. But since the very notion of representation is commonly taken to entail the possibility of misrepresentation, higher-order theories seem to be incompatible with this principle: the qualitative properties instantiated by a mental state M are not necessarily identical with the qualitative properties consciously experienced when a certain higher-order state represents M,

because the latter properties may result from the misrepresentation of the former. The higher-order theorist can repeat the same line of reasoning used by Rosenthal to deal with the interpretation of the objection from intimacy in terms of immediacy: the subjective appearance of infallibility does not entail actual infallibility, but only the impossibility of recognizing (within conscious experience) when one's higher order representations misrepresent their target. Thus, the only mandatory task for the higher-order theorist is to explain why subjects cannot notice in their experiences instances of the relevant appearance-reality distinction. And since it is how the first-order state is represented that fixes the way in which the subject will experience that state, the subject will never be able to notice, within phenomenology, that such states may have qualitative properties different from the ones that her experience ascribes to them. That is, the intimacy of the connection between subjects and experiences may be captured by construing the notion of inner awareness in terms of *constituting* representation (Kriegel 2009, 109): the relevant higher-order representations may not only determine the existence conditions of phenomenal character (by endowing subjects with inner awareness) but also fix its identity conditions (by constituting the qualitative character of conscious states). And if "the first-order state can contribute nothing to phenomenology apart from the way we're conscious of it" (Rosenthal 2004, 32), then no subject could ever notice, within phenomenology, that her mental states have properties different from the ones that her experience ascribes to them: it would subjectively seem that, indeed, the inner awareness of one's conscious states does not admit the possibility of an appearance-reality distinction – despite the possibility of misrepresentation. Hence, the higher-order theorist may conclude that it is "perfectly coherent to suppose that a mental state may represent itself [or another state] to be a certain way when in reality it is not that way" (Kriegel 2009, 136), without renouncing acknowledge the intimacy datum.

However, this interpretation of the privileged access to one's own conscious states seems to build into higher-order theories a further controversial commitment: once the 'constituting representation' view of inner awareness is assumed, it seems that the higher-order theorist should also allow for the possible non-existence of the mental state made conscious, i.e., for the possibility of targetless higher-order representations giving rise to conscious states (Neander 1998, 420). In fact, Rosenthal acknowledges that "a case in which one has a HOT along with a mental state it is about may be subjectively

indistinguishable from a case in which the HOT occurs but not the mental state”, while also conceding that “having a HOT cannot of course result in a mental state’s being conscious if that mental state does not even exist” (1997, 744). By contrast, Kriegel argues that such a possibility is logically incompatible with self-representationalism, as “it is incoherent to suppose that a mental state may represent itself to exist when in reality it does not exist”, for “if it does not exist, it cannot represent itself—it cannot represent anything” (2009, 136), and concludes that this fact provides evidence for the superiority of intrinsic theories, insofar as they can preserve the “obvious truism” that “for any subject S and time *t*, there is something it is like to be S at *t* iff there is a mental state M, such that (i) S is in M at *t* and (ii) M is conscious at *t*” (Kriegel 2009, 130).

A similar criticism of extrinsic theories is put forward by Block (2011a, 424-5), who argues that the possibility of targetless higher-order representations giving rise to conscious states exposes a structural incoherence of those views. The charge of inconsistency rests on the observation that the possibility of targetlessness shows how the necessary condition for the instantiation of conscious first-order states assumed by extrinsic theorist, i.e., their being the object of a higher-order representation, is incompatible with the thesis that the instantiation of a higher-order representation is sufficient for entertaining conscious states – since the sufficient condition can be satisfied without also satisfying the necessary condition (because there can be a conscious state without anything being the object of an appropriate higher-order representation). However, it seems that the extrinsic theorist need not be committed to the necessary condition stated above: even though any unconscious mental state must be the object of a higher-order representation in order to become conscious, it does not follow that any intentional object of higher-order representations must be an existent object – i.e., that the relevant higher-order representations cannot be about non-existent mental states, making the subject conscious of them even though they are not actually instantiated (Coleman 2015, 2708). That is, the notion of ‘being in a conscious state’ may be interpreted in intensional rather than extensional terms, i.e., in such a way that “what counts for somebody’s being in a conscious state is just the occurrence in one’s stream of consciousness of the relevant subjective appearance, the appearance of being in the state in question” (Rosenthal 2011, 432). Accordingly, although the possibility of being conscious of non-existent mental states may appear as counterintuitive, the higher-order

theorist can reject the charge of inconsistency by simply holding that “being in a conscious state is not being in that state and being conscious of being in it, but simply being conscious of oneself as being in the state” (Rosenthal 2004, 41). In fact, this conception of conscious states seems to directly follow from the assumption of the constituting representation view: if “the first-order state can contribute nothing to phenomenology apart from the way we’re conscious of it” (Rosenthal 2004, 32), why should one suppose that such a state must exist to be represented?

As mentioned above, Kriegel rejects this possibility by appealing to the intuitive idea that a mental state cannot represent itself unless it is an existent state. However, given his mereological conception of conscious states’ structure and his commitment to the constituting representation view – required to capture the intimacy datum – it is natural to wonder whether the presence of the first-order state is in fact necessary for the higher-order representation to give rise to consciousness (Coleman 2015, 2715-6). After all, it is the higher-order representation alone – a mental state distinct from the first-order state made conscious (though allegedly unified with it into a conscious state) – that is supposed to be responsible for the constitution of phenomenal character. Kriegel justifies the necessity of the unification of the first-order and higher-order state in a mereological complex (rather than a mere sum) by appealing to the experienced subjective givenness of the first-order state’s phenomenal qualities: in “a perceptual experience of the blue sky, the perception of blue and the awareness of that perception are unified by some psychologically real relation whose dissolution would entail the destruction of the experience” (Kriegel 2009, 222). But given that those experienced phenomenal qualities are only experienced *qua* contents of the higher-order representation, as the “properties [of first-order states] are not part of the experience’s phenomenal character, indeed are not phenomenologically manifest in any way” (Kriegel 2009, 110), it is not clear why those properties should be necessarily carried by an existent first-order state (unified with its representation) in order to appear in conscious experience.

The self-representationalist may attempt to reject the possibility of targetlessness by insisting that, although the properties of first-order states are not the qualities that appear in conscious experience, the unity of the first-order state and its higher-order representation in a mereological complex is required to ground the indirect self-representation necessary for the constitution of subjective character (Coleman 2015,

2714). Yet, it seems that indirect self-representation could obtain even in the absence of the first-order state: given a lone higher-order state representing a non-existent mental state, the higher-order representation may indirectly represent the whole state of which it is part by directly representing its non-existent first-order intentional object. For, on the one hand, it seems that instances of ordinary (i.e., non-reflexive) indirect representation do not require the existence of the directly represented object: just like a painting could indirectly represent an entire house by directly representing only its front even if the directly represented item were demolished (or never existed), so too a higher-order representation could indirectly represent the whole conscious state of which it is part by directly representing a first-order state even if that first-order state were to go out of existence (or never existed in the first place). And, on the other hand, it seems that the possible non-existence of the directly represented item should not be precluded in the case of consciousness just because of the addition of reflexivity to the indirect representation. For even though the painting alone doesn't represent itself, "plausibly all that's needed [for self-representation] is for the painting to be inside the house it depicts", and yet "painting and house are not a complex", for "no entity is destroyed if the painting is removed" and "if the house burns down, but the painting survives, the painting can still indirectly self-represent by representing its old home" (Coleman 2015, 2715, fn. 31).

The self-representationalist may object that, at least in some cases, the possibility of representing a whole via the representation of a part of it seems to depend on the existence of the whole: in the case of the painting, for example, the difference-maker between the painting only representing a house-front and the painting also indirectly representing a whole house (by representing its front) may be precisely whether or not the house really exists: if there is only a front, then the painting will not indirectly represent anything; if there is a whole house, then the painting will indirectly represent a whole house (and possibly also itself, as a part of it) – and if neither the front nor the house exist, then the painting may indirectly represent a fictional house (and thus also possibly itself as a fictional part of it) in virtue of representing its fictional front, but it could not indirectly represent a real house (nor represent itself as a part of a real house).<sup>80</sup> Analogously, it may seem likely that a higher-order representation could only indirectly represent itself by

---

<sup>80</sup> Thanks to Tom McClelland for raising this concern.



representing some first-order states that do in fact belong to a greater whole, including the relevant (allegedly) indirectly represented higher-order properties – for, otherwise, it would be as if a painting could represent itself in virtue of indirectly representing an existent house (of which it is part) that lacks a front by representing a fictional front. However, independently of the exact criteria for indirectly representation,<sup>81</sup> it is doubtful that the analogy with painting should be taken this far, as in the case of self-representationalism the relevant higher-order representation plays both the role of the painting and the role of the house (minus its front – which, in the analogy, corresponds to the represented first-order state). That is, unlike the painting, the only (arguably) possibly non-existent item is the directly represented one (i.e., the first-order state), for the rest of the (allegedly) complex conscious state consists in the higher-order representation that, moreover, is wholly responsible for the identity of the phenomenally experienced contents. Thus, ultimately, we must ask the self-representationalist whether there is any reason to believe that the directly represented first-order state one is conscious of needs to be (or at least to have been at some point in time) properly integrated with its representor. According to Coleman, the answer is negative:

The HO component is ‘none the wiser’ about the non-existence of its intentional object, and this represented object is putatively integrated with it in the way an existent sensory component would be [...].

In our targetless case it’s the constitutive representing of [...] qualities (as ascribed to the sensory component) that sees them enter the subject’s stream of consciousness. Now the sensory component is gone, such representation persists. The qualities that now figure in consciousness are the same, and their source is the same; namely, the aforementioned constitutive HO representation (2015, 2715-6).

In a similar spirit, Rosenthal argues that even granting that self-representationalism gives us “a nonarbitrary reason to say that the target is never absent” (2004, 32), the very possibility of higher-order misrepresentation, due to the self-representationalist’s

---

<sup>81</sup> Which, in any case, are unlikely to involve such a tight connection between the possibility of indirect representation and the existence of the indirectly represented whole. For example, if someone finds an old drawing, made by him when he was a child, depicting the front of his old home – that has been demolished in the meanwhile – it is unlikely that such a drawing will represent for him only the front of his old home just because the indirectly represented whole went out of existence.

commitment to the constituting representation view, is sufficient to pave the road for instances of targetless higher-order representation:

The distinction between an absent target and a misrepresented target is in an important way arbitrary. Suppose my higher-order awareness is of a state with property P, but the target isn't P, but rather Q. We could say that the higher-order awareness misrepresents the target, but we could equally well say that it's an awareness of a state that doesn't occur. The more dramatic the misrepresentation, the greater the temptation to say the target is absent; but it's plainly open in any such case to say either. The two kinds of case, moreover, should occasion the same kinds of phenomenological perplexities, if any. A higher-order awareness of a P state without any P state would be subjectively the same whether or not a Q state occurs. The first-order state can contribute nothing to phenomenology apart from the way we're conscious of it. [...] And, since misrepresentation would occasion the same difficulties, if any, that would occur for an absent target, an intrinsic theory has no advantage in this connection over a theory on which the higher-order awareness is distinct from its target (2004, 32).

Kriegel considers Rosenthal's challenge and tries to address it by holding that there is "very certainly a substantive distinction between targetless higher-order representations and higher-order misrepresentations of lower-order states' properties", insofar as "even if there happens to be an unrelated lower-order state when someone is in a higher-order state, there are facts of the matter pertaining to what makes the latter represent something" and "we cannot just 'interpret' it to represent the former in the absence of any relationship between them" (2009, 135). Yet, even accepting that the blurriness of the divide between higher-order misrepresentation and targetless higher-order representations is not an unsurmountable obstacle for the self-representationalist, it should be remembered that the allegedly consciousness-supporting relationship between higher-order and first-order representations is characterized by Kriegel in phenomenological terms – such that, e.g., "when I have a perceptual experience of the blue sky, the perception of blue and the awareness of that perception are unified by some psychologically real relation" (2009, 222), and this 'real relation' is in turn understood in terms of neural synchronization. But since the suitably synchronized "first-order state can contribute nothing to phenomenology apart from the way we're conscious of it" (Rosenthal 2004, 32), it is not clear why the mechanisms responsible for the physical relationship between the realizers of first-order and higher-order representations should be sensitive to the 'facts of the

matter' that determine whether the latter actually represent the former (or, at least, it is not clear why those mechanisms should be infallible in deciding whether an erroneous higher-order representation counts as a misrepresentation or as a targetless case). Ultimately, just like the phenomenological thesis that we consciously experience the unity of first-order and higher-order representations does not imply anything more than the fact that those higher-order representations represent themselves as being unified with first-order representations – which are experienced as higher-order represented contents that may not have an existent referent – the empirical thesis that consciousness depends on the neural synchronization of first-order and higher-order states does not guarantee that the represented first-order content of a higher-order state is an accurate enough representation of the first-order state with which is synchronized to effectively count as having an existent referent.

Therefore, despite Kriegel's affection for the "obvious truism" that being in a conscious state means being in that state *and* being conscious of being in it, the essential connection between representationalist higher-order theories and the constitutive representation view leads intrinsic theorists to the same conclusion reached by extrinsic theorists: representationalism about inner awareness requires the higher-order theorist to "retreat on the claim that a state's being conscious is strictly speaking relational" (Rosenthal 2005, 179), i.e., to reject the idea that inner awareness essentially involves a relation between the first-order state made conscious and its higher-order representation.

Clearly, there is a sense in which both views retain some degree of 'relationality'. For according to the extrinsic theorist, the content of the higher-order state represents a relation between the subject and the (possibly non-existent) experienced first-order content, and according to the intrinsic theorist the higher-order representation establishes a relation intrinsic to the conscious state, involving its qualitative and its subjective character. But no actual relation with the first-order state allegedly made conscious is required. Thus, although both views set out to explain creature-consciousness in terms of state-consciousness, they end up explaining the former only in terms of the *appearance* of the latter, as no first-order state is actually required to be state-conscious in order to give rise to conscious experience. Though one may accept that in the 'good' cases in which the higher-order representation is accurate, there will be state-consciousness as well, this is simply not sufficient to grant that the first-order state made conscious has any

role in the constitution of consciousness. In what follows, it will be argued that although this fact does not entail the incoherence of higher-order theories (as suggested by Block), it naturally leads to conceiving consciousness as an illusion – thereby taking away the only clear advantage of intrinsic theories over extrinsic theories.

#### 4.3.2. *Representationalism and illusionism*

Although the appeal to the notion of constituting representation allows the higher-order theorist to address the objection from intimacy and capture the epistemically privileged status of the objects of one's inner awareness, it comes at a high cost. Since first-order states have no role in determining the qualitative character of conscious experience, it follows that "the subject simply *never becomes conscious of the first-order state at all*, even in the good case" (Coleman 2015, 2709). That is, if the properties of first-order states "are not part of the experience's phenomenal character, indeed are not phenomenologically manifest in any way" (Kriegel 2009, 110), then we may at best "consciously experience veridical echoes of sensory states; but of the sensory states we are not conscious": the constituting higher-order representations necessarily "block anything else getting in" the conscious experience except for the qualities those representations attribute to the first-order states we are not conscious of (Coleman 2015, 2710). Thus, rather than providing an explanation of what makes our mental states conscious – what makes us aware of our actual mental life – higher-order theories can only explain the conditions required to give us the *impression* that some of our mental states are conscious, as "what counts for somebody's being in a conscious state is just the occurrence in one's stream of consciousness of the relevant subjective appearance, the appearance of being in the state in question" (Rosenthal 2011, 432). This feature of higher-order representationalism directly leads to illusionism about consciousness, i.e., the view that although experiences are real, phenomenal consciousness is an illusion, and that the hard problem can be dissolved by solving the meta-problem – for all there is to explain about consciousness is why it seems to us that we have conscious states endowed with phenomenal properties, rather than why those properties really exist.

The higher-order representationalist may try to reject illusionism by appealing to the analogy between our relation with our own mental states and our perceptual relational with external objects: just as we generally reject the sceptics' worries and believe that we are usually *really* conscious of external objects even though we could misrepresent or

hallucinate them, we should believe that we are really conscious of our mental states even though they do not directly appear in conscious experience. That is, just as the traditional problem of illusion is generally taken to only imply that we are indirectly presented with existent external objects, and not that we live in a simulation in which no represented object of perception really exists, the problem of illusion that stems from higher-order representationalism does not provide substantial reasons to reject the claim that we really are conscious of the represented first-order states and their properties (when they exist and are not radically misrepresented).

Yet, as suggested by the intimacy datum, there seem to be essential differences between our representational relation with the world and our relationship with our conscious mental states. In the case of perception, despite the impossibility of arriving at a definitive refutation of the sceptics' worries, we have at least a variety of ways to put our fallible representations to test and assess their accuracy (e.g., comparing the information from different sensory modalities, or constructing devices that make up for some of our biases) that provide us with some degree of confidence in the existence of most of our representations' referents. By contrast, in the case of conscious experience, the only reason to suppose that our first-order states really are *as they are experienced* is provided by inner awareness.<sup>82</sup> But if it is inner awareness that determines those states' appearances and the properties of first-order states never make it to consciousness, it seems that we have no substantial reason to suppose that those states really have the phenomenal qualities they seem to have within phenomenology (assuming a realist attitude towards the hard problem).

Still, the higher-order representationalist may try to deny his commitment to illusionism by appealing to the absence of an appearance-reality gap in consciousness – not only in epistemic terms but also in metaphysical terms. That is, perhaps phenomenal consciousness essentially *is the impression* that a mental state has certain qualities, and for that state to have a phenomenal character is simply to be subjectively presented as having those qualities. And if the impression is real, then phenomenal consciousness is real as well: it turns out that higher-order representationalism does not presuppose that

---

<sup>82</sup> Clearly, it is possible to question this thesis – e.g., by appealing to Rosenthal's (2005) Quality-Space theory. However, this move would require one to deny the existence of intrinsic differences between conscious and unconscious qualities, which would directly lead, within the framework of higher-order representationalism, to refusing to take the hard problem seriously.

phenomenal consciousness is an illusion, but only that consciousness is a matter of how things appear to a subject. Thus, even though illusionism is compatible with higher-order representationalism – since the higher-order theorist appealing to constituting higher-order representation may deny that our mental states have phenomenal character and instead “focus on explaining why they seem to have them” (Frankish 2016, 17) – the higher-order theorist may accept the reality of phenomenal character by conceiving a mental state’s having phenomenal character in terms of that state’s seeming to have one.

Yet, the difference between the illusionist and the representationalist readings of the idea that consciousness is just a subjective impression seems to be purely verbal: how could one find a substantial difference between the illusionist thesis that conscious experience is the illusion that mental states have phenomenal characters and the representationalist thesis that conscious experience has an illusory content, in that it appears as presenting us with our first-order states and their qualities but (given the constituting representation view) it does not? Just like the illusionist, the higher-order representationalist holds that conscious experiences are real, but phenomenology misleads us – as it presents us with mental states that are never phenomenologically manifest. And if all that is required for a mental state to have a phenomenal character is subjectively appearing as having one, then there is no theoretical need to admit into our ontology something like phenomenal properties, i.e., the properties of first-order states (at least partly) responsible for the constitution of those states’ phenomenal characters (in virtue of which there is something it is like to be in those states). Thus, if the instantiation of phenomenal properties by first-order states simply amounts to the subjective appearance that those properties are instantiated, it seems that there is no substantial reason to introduce those properties into our ontology in the first place, and to reject the more parsimonious ontology put forward by illusionism: higher-order representationalism does not address the question of how intrinsic phenomenal properties of conscious states can come into existence – rather, it is only concerned with the illusion problem, i.e., the challenge of explaining how illusions of phenomenality can arise in non-phenomenal systems.

Finally, the higher-order representationalist may recognize that no intrinsic phenomenal properties should be ascribed to conscious first-order states and, yet, still maintain that he can take the hard problem at face value. Perhaps the difference between

illusionism and higher-order representationalism is in fact purely verbal, but the kind of illusionism under examination does not really comport the dissolution of the hard problem – as it does not involve the strong illusionist claims that consciousness does not exist and that intuitions about the presence of a hard problem only arise in virtue of mistaken introspective attributions of phenomenal properties to conscious states, but only the claim that conscious states subjectively appear as different from how they really are:

Illusionists who want to use illusionism to dissolve the hard problem of consciousness should be strong illusionists. [...] The basic reason, as I see it, is that the hard problem does not turn on the claim that consciousness is intrinsic, or non-physical, or non-representational, or primitive, and so on. For example, we can be agnostic about whether consciousness is intrinsic, or hold that it is extrinsic, and the hard problem arises as strongly as ever: why is it that when certain brain processes occur, there is something it is like to be us? (Chalmers 2018, 49).

However, it seems that taking weak illusionism not to be an attempt to dissolve the hard problem leads us to ignore its deflationary spirit. Clearly, the weak illusionist is not committed to the strong illusionist claim that “cognitive scientists should treat phenomenological reports as fictions – albeit ones that provide clues as to what is actually occurring in the brain” (Frankish 2016, 26), as those reports (that constitute the basis for the intuition that there is in fact a hard problem) are based on some real phenomenon that gives rise to the appearance of phenomenal properties in conscious experience (and not only in introspection). Thus, there is a sense in which, rather than dissolving the hard problem, the weak illusionist accepts the reality of the hard problem and proposes a solution to it. However, although the weak illusionist can be seen as “taking consciousness seriously”, he also appears to attempt “to redefine the phenomenon in need of explanation as something it is not” (Chalmers 1996, X). For although he grants that we should explain why there is something it is like to be a subject, he also denies that there is something it is like for a subject *to be in* a certain first-order mental state – as there is only something it is like to have the (non-relational) impression of being in that state. That is, considering weak illusionism to be a genuine way of taking the hard problem at face value leads to overlooking the fact that, when it comes to establishing why we feel like there is in fact a hard problem, the weak illusionist shares the strong illusionist thesis that “our sense that it is *like something* to undergo conscious experiences is due to the fact that we systematically misrepresent them (*or, on some versions, their objects*) as

having phenomenal properties” (Frankish 2016, 13; second emphasis mine).<sup>83</sup> The only difference between the two is that the strong illusionist believes that the so-called hard problem is not that hard because we misrepresent conscious experiences as phenomenal in introspection, whereas the weak illusionist holds that the hard problem is not so hard because we misrepresent the objects of conscious experiences (i.e., first-order states) as phenomenal even though the properties of those mental state are never really experienced.

The appeal of self-representationalism is precisely that, unlike extrinsic theories, it allegedly allows the higher-order theorist to reject this illusionist claim and hold that the objects of conscious experience have in fact phenomenal properties – in virtue of their integration with higher-order representations supposedly essential for the constitution of conscious states. But once it is recognized that such integration is in fact unnecessary for the constitution of consciousness, it seems that assuming the intrinsic appearance of the phenomenal properties of first-order states can only lead us to conclude that those properties are simply the objects of the subjective illusion we call consciousness. That is, just like the extrinsic theorist, the self-representationalist ends up conceiving of the intimacy of our relationship with our conscious states, i.e., the fact that phenomenal appearance seemingly collapses onto phenomenal reality, as only due to appearances determining phenomenal reality. And, ultimately, the only good reason to assume the reality of the phenomenal properties of the first-order states we are conscious of is to endorse the opposite thesis (i.e., that it is phenomenal reality that determines appearances), which is inconsistent with the (constitutive) representationalist conception of conscious experiences as subjectively presented collections of qualities fundamentally disconnected from the mental states to which we ordinarily ascribe them. Therefore, it seems that higher-order intentionalism may be reconciled with a non-deflationary approach to the hard problem only by dropping the conception of inner awareness as a representation of first-order states. The purpose of the following chapter is to outline the main strategies to develop such a view (without giving up naturalism).

---

<sup>83</sup> The higher-order representationalist may object that he does not hold that we misrepresent first-order states as having phenomenal character but, rather, that those states get to have a phenomenal character because we represent ourselves as being in those states – so that no representation of phenomenal properties of states is required. Yet, it is not clear how higher-order representations could fix the identity conditions of phenomenal character of without representing first-order states as having phenomenal qualities.



## 5. Higher-Order Intentionalism and Realism about the Hard Problem

The purpose of this chapter is to present the explanatory strategies available to the higher-order theorist unwilling to commit to the illusionist outcomes of representationalism. The possibility of developing these alternative varieties of higher-order intentionalism follows from a simple observation: even though the terms ‘intentionalism’ and ‘representationalism’ are often treated as synonyms, there is a significant conceptual difference between the claim that inner awareness is constituted by higher-order intentionality and the claim that inner awareness is constituted by a higher-order representation of the first-order state made conscious. The idea that the instantiation of intentional properties (i.e., the presence of ‘aboutness’) does not obviously entail the instantiation of a representation of the object those properties are about can be introduced by analogy, considering the peculiar ontology that characterizes so-called ‘display sentences’, such as “the sign ‘Keep Off’ on a road; ‘Shake well’ on a bottle; the date written at the head of a letter; ‘New, Improved’ on a cereal box; ‘\$100’ on a dress, etc.” (Zemach 1985, 195). In all these examples, it is apparent that the subject matter of the sentence is not represented by ink marks but rather, being within reach, it is simply presented – embedded in the discourse without using a symbol that stands for it.<sup>84</sup> For instance, when reading ‘Shake well’ on a bottle, it is quite intuitive that the sentence means ‘Shake *this bottle* well’ despite the absence of linguistic symbols directly representing the bottle itself, in virtue of the particular context in which that sentence is tokened (i.e., in virtue of its being physically attached to the object its subject matter refers to). This phenomenon may be interpreted within a representationalist framework, but it need not.

An eminent example of a representationalist interpretation of display sentences – applied to the debate under consideration – comes from Kriegel (2009). He points out that in reading the words ‘under construction’ painted on a bridge, those words do not account for the whole propositional content tokened, as “the bridge itself functions as the subject term in the *sentence* that vehicles the proposition that makes up your thought’s content” (2009, 163). Then, he suggests that the bridge enters the content of that sentence by

---

<sup>84</sup> Zemach (1985) attributes the idea to Searle (1969).

representing itself and that, analogously, mental states enter the contents of consciousness in virtue of their self-representational properties; such that, e.g., to have a thought the content “this very thought is being thought by the mess-maker” consists in having “a certain internal occurrence or token with the Mentalese ‘being thought by the mess-maker’ so to speak ‘posted’ on it” (2009, 164).

However, this self-representational interpretation of display sentences may be rejected by the higher-order theorist wanting to avoid the pitfalls of illusionism, as (e.g.) the bridge may enter the content of a sentence without performing any representational role: perhaps “the bridge is *present* in the sentence, allowing the sentence overall to say something about it” (Coleman 2015, 2716). And, similarly, first-order states may become intentional objects of inner awareness without being represented in virtue of being cognitively ‘within reach’ (rather than being made cognitively available in consciousness in virtue of being represented). That is, while Kriegel is committed to interpreting the ‘subject matter’ of inner awareness (i.e., conscious states) in self-representational terms – by characterizing it as being constituted by the higher-order representation of a first-order state that indirectly represents itself by directly representing the (possibly non-existent) first-order part of itself – it seems possible for the higher-order theorist to reject this interpretation of mental display sentences and, instead, conceive of consciousness as involving the direct presentation (rather than the representation) of the mental states made conscious.

In what follows, it will be argued that this alternative interpretation of the analogy between consciousness and display sentences can allow for the development of extrinsic higher-order theories able to take the hard problem at face value – either within the framework of the state view, by adopting a modest extrinsic view while holding that only the qualitative character of first-order states is displayed in consciousness (§5.1), or within the framework of the subject view, by adopting a modest or an ambitious extrinsic view while taking subjective character to be displayed in consciousness as well, though as an extrinsic aspect of the phenomenal character of the first-order states made conscious (§5.2).

## 5.1. Quotational Higher-Order Thoughts

According to the Quotational Higher-Order Thought (QHOT) theory of consciousness (Coleman 2015; 2017; 2018), first-order mental states are made conscious by higher-order thoughts able to “‘quote’ a sensory state, forming a larger composite structure wherein the sensory state is displayed” (Coleman 2015, 2717).<sup>85</sup> Thus, the fundamental principle of the QHOT theory is that first-order states, not their representations, are the essential constituents of conscious experience – as they can be directly presented (i.e., exhibited) in consciousness:

When a first-order state is quoted in the requisite way, the result is a mental display sentence, which supports a conscious state, on the analysis. [...] In QHOT theory mental state tokens are recruited for display in HO quotational structures that supply consciousness of said first-order content. [...] QHOT theory’s quotational structures are nonrepresentational: for there is no need to represent a token state which is actually present. (Coleman 2018, 43).

The idea that first-order states can partly constitute the content of a larger mental state token is reminiscent of the quotational account of phenomenal concepts (Papineau 2002; 2006; Balog 2012), according to which concepts referring to conscious experiences are formed by means of a cognitive “mechanism that operates on an experience and turns it into a phenomenal concept that refers to either the token experience, or to a type of phenomenal experience that the token exemplifies” (Balog 2012, 33). However, while on the quotational account of phenomenal concepts a quoted mental state token is already a conscious item and is typically used to represent in thought the phenomenal type it belongs to, on the QHOT theory “the quoted elements are not yet experiences”, as the appeal to mental quotation serves precisely the purpose of “explaining what turns [e.g.] sensory states *into* experiences” (Coleman 2015, 2718). Moreover, for the same reason, the kind of mental quotation involved in the QHOT theory is not supposed to provide a representation of what the QHOT is about, and thus essentially involves the token first-order state itself rather than the type it belongs to. Accordingly, “the quotational higher-order thoughts that supply consciousness are envisaged as very *thin*, best modelled as

---

<sup>85</sup> The idea that consciousness involves higher-order quotation has also been suggested by Picciuto (2011). However, he characterizes the relevant quotational structures as phenomenal concepts that work as “demonstratives that necessarily display their referents” (2011, 132), and therefore – even granting that demonstrative reference can constitute rather than presuppose consciousness (cf. §4.1.1) – cannot avoid the illusionist outcomes of representationalism.

demonstrating ‘frames’, with a ‘slot’ for the sensory state” (Coleman 2015, 2718), which may be characterized as follows:

This state is present: “ \_\_\_\_\_ ”

This feature of QHOTs allows the higher-order theorist to avoid the illusionist outcomes of higher-order representationalism, by making it possible to capture the intimacy of the connection between subject and experience without construing the notion of inner awareness in terms of constituting representation: on the QHOT theory, the consciousness-generating higher-order states partly determine the existence conditions of phenomenal character (by endowing subjects with inner awareness) without fixing its identity – as they do not constitute the exhibited qualitative character. Thus, higher-order misrepresentation simply cannot occur, as the phenomenal contents of experience are entirely provided by the first-order states made conscious, which are directly displayed. The same point can be expressed by analogy. Within the linguistic domain, different types of quoted contents are associated with different degrees of possible misrepresentation:

If I want quotationally to represent what Florence said yesterday in the heat of argument I can say ‘She said: “Get out and never come back”’. A little closer to the QHOT model, I can play a tape-recording of what she said (was I sufficiently self-possessed to make one), saying ‘She said this: \*click\*’, i.e. playing the tape recording at the relevant point. Closer still, I could summon Florence to repeat what she said, saying ‘She said: “\_\_\_\_\_”’, and then letting her rip. I am not yet employing her *very utterance*, though, so we can imagine one further, somewhat outlandish, case. Had we a time machine, we could return to the instant Florence was about to shout at me, and I could say (observing my wretched past self): ‘She said “\_\_\_\_\_”’, indicating the token utterance (Coleman 2018, 44).

That is, the possibility of misrepresenting reality by means of linguistic quotation is available just in case the quoted item is represented, rather than displayed. While verbally repeating what Florence said obviously leaves the speaker with a significant freedom to misrepresent her utterance – even if it is Florence herself that repeats what she said (as she could repeat it incorrectly), and similarly using a tape-recording device does not guarantee truthfulness since the tape could be tampered with, if the quotational device embeds the quoted content in the very same moment in which it is tokened, no misrepresentation can occur, because that content “is no longer *represented*, but

*exhibited*”, i.e., “her utterance is used by me to present itself” (Coleman 2018, 44). Analogously, the possibility of misrepresentation of the first-order states made conscious by QHOT is not allowed, “for what supplies the sensory content of the quotational conscious state is just *that* very sensory state itself, with *its* content” (Coleman 2015, 2721). And although it may be possible to envisage instances of misrepresentation deriving from the part of the QHOT’s content that is not the quoted first-order state, e.g., if the QHOT’s part asserting ‘this state is present’ were substituted with “the qualitative *opposite* of this state is here”, it is still the case that “sheer quotation, a higher-order operation, cannot amend the quoted” (Coleman 2015, 2723).<sup>86</sup>

Moreover, for the very same reason, the possibility of targetless consciousness-generating higher-order states seems to be precluded as well. For, even though the instantiation of a targetless QHOT seems at least conceivable, given that the higher-order quotational element does not determine qualitative character it is natural to suppose that, just like “without the quotational frame, a sensory state is not displayed, so by hypothesis fails to be the object of awareness”, in the same way “a lone HO state cannot support an experience, since it proffers no sensory content – it simply lacks the resources” to give rise to a conscious experience (Coleman 2015, 2718). That is, since the QHOT is supposed to constitute inner awareness but not what one is made aware of, it follows that “if a QHOT has no sensory target to embed, it could at most arouse subjective awareness of (a state of) *nothing*” (Coleman 2015, 2722). Determining whether or not such a state of awareness would genuinely result in a conscious experience goes beyond present purposes. According to Coleman, it would not, because “an experience literally of nothing is simply no experience”, as “subjective awareness must be (intentionally) *of something*” (2015, 2722). However, even granting that it is possible to experience nothingness by having an empty QHOT, e.g., through meditation,<sup>87</sup> such an experience would be immediately distinguishable from any other type of experience due to its peculiar lack of qualitative character. For the possibility of targetless consciousness-generating higher-order states poses a problem to the higher-order theorist (wanting to reject illusionism) precisely because it exposes the irrelevance of first-order states in the constitution of the

---

<sup>86</sup> Coleman suggests that “the most that could happen is that the subject acquires an erroneous (and, if conscious, discomfiting) *belief* that what is an occurrent conscious state is distorted in consciousness” (2015, 2723).

<sup>87</sup> Coleman proposes to challenge the description of alleged experiences of nothingness achieved through meditation as involving “an exceptionally general, diffuse, sense of oneself, or the universe at large” (2015, 2722, fn. 52).

qualitative character of conscious experience – thereby leading to the conclusion that “the subject simply *never becomes conscious of the first-order state at all*, even in the good case” (Coleman 2015, 2709). But even granting, against Coleman, that a targetless QHOT can produce *one* peculiar non-qualitative kind of conscious experience does not lead to the same conclusion, as it does not lead to accepting that a non-empty QHOT could give rise to an indistinguishable conscious experience (in which the presence of the quoted first-order state would be irrelevant). That is, even if it is possible to experience *nothing* in virtue of having targetless QHOTs, this possibility does not open the door to illusionism as the possibility of experiencing *something* in virtue of having targetless higher-order representations does, because no qualitative character apparently had by first-order states would be involved in the former type of experience.<sup>88</sup>

However, the higher-order representationalist may object that the possibility of capturing the intimacy of the relation between subject and experience by appealing to mental quotation – instead of merely accounting for the subjective impression of intimacy by appealing to constituting representations – is not a ‘truth-indicating’ virtue of QHOT theory but only a ‘desire-satisfying’ virtue, such that it could only make the theory appear preferable to its competitors other things being equal. Then, the higher-order representationalist could argue that other things are not in fact equal, because the notion of representation is already well-established and used to explain a multitude of mental phenomena, whereas the notion of non-representational mental quotation is posited exclusively in order to account for consciousness, and it is not wholly clear what the cognitive mechanism responsible for its existence could consist of. That is, even granting that it would be in fact preferable to construe consciousness in such a way that it delivers genuine intimacy with one’s own conscious states rather than only its appearance, it is still the case that “quotation operates on a token of what is quoted, and if the token is a property of a distinct state, machinery would have to be posited of a wholly unclear type”

---

<sup>88</sup> There is a sense in which “nothing” can be conceived as being “something” (given a non-Quinean interpretation of the particular quantifier), i.e., nothing could be a non-existent intentional object, defined as “the mereological sum of the empty set”, with which “one can have direct phenomenological acquaintance” (Priest 2014, 56). However, being conscious of nothing in virtue of having an empty QHOT would likely give rise to a noticeably different conscious experience than, e.g., consciously thinking about such a non-existent intentional object, despite the two sharing a lack of qualitative character (on the assumption that cognitive phenomenology, if there is such a thing – does not involve distinctive qualitative properties irreducible to sensory qualities (Voltolini 2016)). For the latter experience would involve the awareness of a first-order state absent in the former experience: i.e., an empty QHOT would give rise to an experience of nothingness, while a QHOT embedding a thought about nothing-the-intentional-object would give rise to the experience of thinking about nothingness.

(Rosenthal 2018, 63), resulting in a theory of consciousness significantly less informative than representational theories. Coleman suggests that mental quotation may be conceived as a form of “*proto-quotation*” posited to explain our ability for linguistic quotation in the same way in which “linguistic reference is often considered derivative upon the capacity for mental reference” (2015, 2726). But in order to conclude that one should posit this cognitive mechanism to explain the constitution of inner awareness “we need some independent understanding of how the mental quotation or splicing works”, such that the case for the QHOT theory does not “rest simply on the ability to generate the desired intimacy” and one may suppose, instead, that there is a “reason independent of that to think that the answers [provided by the QHOT theory] fit with a theoretically sound picture of how intentional content and mental qualitative character work” (Rosenthal 2018, 64).

However, Coleman does provide such a reason, by specifying the semantic mechanism involved in mental quotation in terms of part-whole constitution: “QHOTs make one aware of sensory states, and the ‘of’ is indeed that of intentionality” even though it is not that of representation because “something can be an intentional object by being *contained* – literally a content” (Coleman 2015, 2727). Following Feinberg (2000; 2005), he suggests that the brain may be conceived as a ‘nested’ hierarchical system in which “the elements composing the lower levels of the hierarchy are physically combined, or *nested*, within higher levels to create increasingly complex wholes”, rather than as a system in which “the lower and higher levels of the hierarchy are physically independent entities” (Feinberg 2005, 39). On this view, the relation between earlier processing and later cognitive operations on those previously generated contents cannot be cashed out in merely causal-representational terms. For, even though “it is undeniable that neurons cause effects in subsequent neurons as [e.g.] we travel along the visual system in its generation of a visual state”, if those neurons form a nested hierarchical system it follows that “the relationship of these elements to the *final* product is not causal, but is rather compositional”, since “the processing of lower-level neurons does not prompt, but actually helps to make up, the final product, the visual state” (Coleman 2019, 66). The resulting view is that mental states are realized by widely distributed structures implemented across different areas of the brain, rather than being sparsely encoded in smaller and highly localized sets of neurons. For example, if the brain is conceived as a

nested system, it follows that there is no “single ‘grandmother’ cell” that “embodies the representation of an entire face” – such that, once certain features of the environment are recognized as one’s grandmother’s face, the earlier visual processing causally responsible for that complex representation becomes redundant – rather, “the conscious representation of the face of one’s actual grandmother requires contributions from diverse and widely separated brain regions” (Feinberg 2005, 29), i.e., the lower-level contents provided by earlier processing perform a constitutive, not a merely causal role in the production of the resulting complex visual state. Analogously, on the QHOT theory, conscious experiences can be characterized as resulting from first-order states’ being physically nested within, rather than represented by, the higher-order states responsible for the constitution of inner awareness:

The quotational frame directs awareness onto the sensory quality it contains: the sensory quality is in this literal sense a content. To be clear, I am not talking metaphorically: I am talking about a physical embedding, on some level of brain organization, of sensory states within the apparatus of awareness. The proposal is effectively that the sense of ‘content’ normally in play in these discussions, for instance in connection with representational mechanisms posited for thought and perception, is to be cashed out via a physical, spatial sense of ‘content’, instead of via the notion of representation (Coleman 2019, 68).

Therefore, it seems plausible that the QHOT theory can indeed “fit with a theoretically sound picture of how intentional content and mental qualitative character work” (Rosenthal 2018, 64), and that its defence need not depend only on its desired ability to capture the intimacy *datum*.

Moreover, for closely related reasons, the resulting view of consciousness put forward by the QHOT theory also comes with a straightforward answer to the generality problem (i.e., the question of why only mental states are the right kind of entities that can become the object of consciousness) that is unavailable to the higher-order representationalist. As considered earlier (§4.1.3), the higher-order representationalist may address the generality problem by adopting an intrinsic view and maintaining that mental states are essentially different from any other object of awareness due to their unique capability to acquire the intrinsic for-me-ness that is experienced in consciousness, or by adopting an extrinsic view (thereby rejecting that inner awareness involves conscious for-me-ness) and trying to argue that awareness of mental states is essentially different from any other



instance of awareness because it involves the representation of something as a property of oneself as an individual (as opposed to the representation of something as a property of external items or of parts of oneself). Yet, this latter strategy directly leads to rejecting the thesis distinctive of realist approaches to the hard problem, i.e., that conscious states have intrinsic phenomenal properties that non-mental entities cannot possess. For, within the framework of extrinsic higher-order representationalism, conscious states cannot be taken to have intrinsic subjective character (on pain of regress), and even if the first-order states made conscious had in fact intrinsic qualities, those properties would be irrelevant for the constitution of phenomenal character. By contrast, abandoning the representationalist conception of inner awareness allows the higher-order theorist to hold that the intrinsic properties of first-order states play an essential role in the constitution of consciousness, since conscious states are supposed to be formed by the embedding of the states made conscious into the awareness-generating QHOTs. Accordingly, it follows that “rocks [as well as non-mental internal states] cannot form part of the mental complexes needed for conscious states” because “consciousness of a state requires it to be ‘neurologically grabbable’” (Coleman 2015, 2706) in such a way that it can be physically nested within a higher-order mental quotation.<sup>89</sup>

Yet, following Levine (2006), it may still be objected that the appeal to the physical embedding of first-order states into QHOTs cannot genuinely provide an account of consciousness able to solve the hard problem. Levine’s objection targets an analogous constitution-based account of phenomenal concepts – according to which “a phenomenal concept affords [direct inner awareness of] the relevant phenomenal property by containing an instance of that property within it” (2006, 162) – but it can be adapted to the case of QHOT theory.

The phenomenal concepts strategy stems from the idea that the hard problem should be addressed by recognizing the presence of an explanatory gap between phenomenal and physical truths while rejecting the inference from the presence of such a gap to the falsity

---

<sup>89</sup> Although it may be objected that internal states (such as states of the liver) may be physically embedded into QHOTs in the same way in which brain states are, it is likely that believing in such a possibility already means presupposing that bodily states are part of cognitive systems – performing constitutive rather than causal roles in cognition. That is, if internal states ordinarily thought to be non-mental are supposed to be “neurologically grabbable”, then it seems that we would have good reasons to believe that those internal states are in fact mental states.

of physicalism.<sup>90</sup> According to supporters of the phenomenal concepts strategy, the explanatory gap is not generated by the (alleged) ontological *sui generis* status of phenomenal properties, but rather by the uniquely direct cognitive access that the concepts formed by attending to one's phenomenal states endow us with. That is, on the one hand, the phenomenal concepts strategy allows to accept that "what is presented by way of phenomenal concepts is distinct from what is presented by nonphenomenal concepts", since "the properties of experience are [...] cognitively present to us" in such a way that our "epistemic relation to them [...] is more intimate, more substantive, than the kind of relation that obtains between our minds and other items" (Levine 2006, 159). But, on the other hand, the difference between phenomenal and nonphenomenal concepts is not explained by differences in what is presented by those concepts but, rather, by differences in *how* those distinct kinds of concepts present their objects – and thus "even if they in fact pick out the very same properties, we find it cognitively difficult to see how this can be" (Levine 2006, 159). A natural way of understanding this unique epistemic feature of phenomenal concepts involves the appeal to the idea that phenomenal states are physical constituents of the phenomenal concepts referring to them, i.e., that we should account for the cognitive presence of conscious states by interpreting literally the "metaphorical language about 'sticking the object right in there'" that "irresistibly comes to mind" (Levine 2006, 163) when thinking about it. But since "cognitive presence [...] is just that: a *cognitive* relation", Levine argues that "it is not at all clear why, or how, *physical* presence translates into cognitive presence" (2006, 162). That is, even granting that phenomenal states may be physically present within the phenomenal concepts about them, given the possibility of multiple realizability of cognitive properties, it is doubtful that the identity of the physical realizers of phenomenal concepts can explain the cognitive presence they afford us:

We are still owed an account of how physical presence alone is responsible for cognitive presence. That is, how does the presence of the relevant state within the physical implementation of the representation become something of which we are aware? [...] The transition from physical

---

<sup>90</sup> The connection between the existence of the hard problem and the presence of an explanatory gap follows from the observation that "the connection between the [extrinsic] neurological description and our [intrinsic] first-person conception of what it's like seems totally arbitrary", i.e., that there is a "sense of arbitrariness that attends the psychophysical reduction as opposed to the sense of intelligibility that attends other theoretical reductions", such that "one feels that this neurological configuration could just as easily have gone [...] with a state that was like nothing at all for the subject" (Levine 2006, 145).

containment to awareness – the special kind allegedly afforded by phenomenal concepts – is still an inexplicable transition. It is subject to its own explanatory gap, just as much as is the original relation between phenomenal properties and their physical correlates (Levine 2006, 163).

In other words, the supporter of the phenomenal concepts strategy suggests that the explanatory gap between truths about phenomenal properties and truths about the physical properties that allegedly realize them should be explained away by appealing to the direct cognitive access afforded by phenomenal concepts, but in doing so ends up giving rise to another explanatory gap between truths about the physical presence of a phenomenal state within the phenomenal concept about it and truths about the intimate cognitive presence of that state allegedly constituted by such physical presence, i.e., “between implementations of cognitive architecture [realizing phenomenal concepts] and whatever it is about phenomenal concepts – in my terms, that they afford genuine cognitive presence to phenomenal properties – that is responsible for the original explanatory gap” (Levine 2006, 165). And, as Coleman himself recognizes, it seems that this objection to the phenomenal concepts strategy may be raised against the QHOT theory as well, insofar as even if “QHOTs are not concepts, nor representational, and the present model is designed as an analysis of what happens in consciousness, not in thought about consciousness”, one can have the impression that “Levine’s criticism retains bite: just how does QHOT theory *explain* awareness?” (Coleman 2019, 72, fn.72). After all, “both proposals aim to account for immediate awareness of mental contents [i.e., cognitive presence] in terms of these contents’ being physically present in concepts or thoughts about them” (Mihálik 2022, 1436). Thus, just as Levine asks why the physical presence of a phenomenal state within a phenomenal concept should explain the peculiar cognitive presence of that state (giving rise to the explanatory gap originally associated with the hard problem), one may ask why the physical presence of (previously unconscious) first-order states within a quotational higher-order frame should explain the intimate cognitive presence of those mental states to their subject – thereby pointing at the presence of “an explanatory gap between the *physical presence* of a [first-order state’s] quality in a QHOT and the quality’s *cognitive presence* for us” (Mihálik 2022, 1439).

However, the task of bridging such a gap is significantly easier for the supporter of the QHOT theory than for the supporter of the phenomenal concepts strategy. For, while the

phenomenal concepts strategy is devised to explain away the intuition that conscious states involve intrinsic phenomenal properties that escape purely functional characterizations (by appealing to the peculiar cognitive access to conscious states afforded by phenomenal concepts), it is possible to endorse the QHOT theory of consciousness while accepting that there is an explanatory gap (giving rise to the hard problem) because phenomenal consciousness involves the experience of intrinsic qualities that cannot be captured in purely extrinsic terms. This is, in fact, Coleman's (2014, 2016) own position: a form of Russellian Monism (i.e., the view that physical objects possess fundamental, irreducible intrinsic properties), labelled as 'panqualityism' by Chalmers (2016), according to which phenomenal character is made of fundamental organized qualities which can exist as "*unexperienced qualia*", i.e., "properties just like the qualia we experience, only without anyone experiencing them" (Coleman 2017, 249). Accordingly, the QHOT theory need not be interpreted as putting forward a complete solution to the hard problem, for it can be seen as only being an attempt "to provide a formal model of the features of [direct inner awareness], and to make the case for the possibility of a physical implementation", (Coleman 2019, 72, fn.72). The resulting picture of consciousness is the one labelled earlier as a 'modest extrinsic view' (§2), according to which consciousness is the extrinsic property of mental states that unveils their pre-existing qualitative aspects, and phenomenal character is constituted by non-essentially conscious intrinsic qualities that can be made phenomenal by inner awareness precisely in virtue of the intrinsic features of those qualities. And since on this view a theory of consciousness is not *ipso facto* a theory of phenomenal properties – as it is a theory of what allows us to experience the intrinsic qualitativity of the properties of first-order states – it follows that there is no explanatory gap between the physical presence of first-order states' within QHOTs and the cognitive presence of their properties, because the latter is supposed to be explained by the conjunction of the former *and* the intrinsic features of what is physically present within the relevant QHOTs (i.e., first-order states' irreducible qualities).

Mihálik (2022) tries to resist this conclusion by arguing that the combination of QHOT theory and panqualityism does not provide a genuine solution to the hard problem able to bridge the explanatory gap, because we might envisage a kind of phenomenal zombies that are "our exact replicas with respect to both the qualities instantiated and the structural

properties constituting the QHOT mechanism” (1441) for whom, nevertheless, no mental quality is cognitively present (despite their physical presence within the relevant QHOT). However, it seems that this rejoinder could only work if one had already assumed that cognitive presence is built into phenomenology, i.e., that subjective character or for-me-ness is a conscious feature of phenomenal character.<sup>91</sup> Since the formulation of zombie-arguments rests on the conceivable absence of experienced qualities given a certain psychological state that, in our world, involves the presence of conscious qualities, it follows that “any item not associated with a set of [mental] qualities is not a valid target for a zombie argument” (Coleman 2017, 278, fn. 73). Hence, within a panqualityist framework, if cognitive presence is not conceived as a conscious item essentially associated with the intrinsic qualities made conscious by inner awareness, there can be no imaginable contrast between one’s conscious mental life and one’s allegedly zombified replica. And since the QHOT theory of consciousness is naturally conceived as an extrinsic state view, it already presupposes the rejection of the claim that the cognitive presence of conscious qualities figures in phenomenal character. For the relevant QHOT, being distinct from the quoted first-order state, must be characterized as an unconscious part of the cognitive machinery responsible for the constitution of conscious experience – as, otherwise, another level of mental quotation would be needed to explain its being conscious (and so on, *as infinitum*).

The critics of the QHOT theory may still object that, even if panqualityism is true and cognitive presence is not an aspect of phenomenal character, zombies may still be conceivable insofar as for any physical relation between a mental state with unconscious qualia a wider neural system – included the one that realizes the quotational embedding of first-order states – it is conceivable for that physical relation to obtain without those qualia being experienced.<sup>92</sup> That is, even though mental states are constituted by categorical qualia with metaphysical necessity, there is no way to determine analytically that a certain physical or cognitive relation will in fact make the subject conscious of those qualia in every possible world. Now, the supporter of the QHOT theory has two

---

<sup>91</sup> As Mihálik himself points out, cognitive presence and the inner awareness constitutive of phenomenal consciousness (that he labels as ‘strong awareness’) are naturally conceived as “two sides of a single coin, capturing the same cognitive relation from two angles: if I am strongly aware of a quality, the quality is cognitively present for me, and vice versa” (2022, 1439).

<sup>92</sup> Thanks to Tom McClelland for raising this concern.

kinds of response available to this rejoinder. On the one hand, he may bite the bullet and accept that the combination of panqualityism and QHOT theory could still allow for the conceivability of zombies, by conceiving the view as an empirical hypothesis concerning the property that constitutes consciousness in this world (and in nomologically alike possible others), rather than as a metaphysical claim concerning what consciousness must be. Within this kind of approach – notoriously defended by Rosenthal (e.g., 2005) – even granting that it is conceivable that there are worlds in which the intimate cognitive access to one’s mental states afforded by QHOTs does not also constitute phenomenal access to those states’ qualia, the supporter of the QHOT theory may still hold that, at least, the view provides a solution as to how to bridge *one* explanatory gap – a solution that does not require phenomenal qualities to magically appear in virtue of their physical presence in phenomenal concepts, but grounds their phenomenal appearance in a direct cognitive relation between subject and intrinsically qualitative properties. On the other hand, the supporter of the QHOT theory may even argue that the objection misses the target, because the whole purpose of explaining inner awareness in terms of higher-order quotation is to find a way of having (independently qualitative) qualia being *directly displayed as present* in one’s mental life. Thus, perhaps the burden of the proof lies on the critics of the QHOT theory, who should convincingly explain how having a direct cognitive access to the intrinsically qualitative properties of one’s mental states could not yield conscious experience – as the inner awareness obtained through quotational embedding is not the ‘standard’ type of cognitive accessibility that intuitively lends itself to zombie-scenarios. This line of argument could be further pursued by appealing to Kriegel’s idea that the non-functionalizable appearance of consciousness does not entail its being non-functional, insofar it may be due, instead, to a peculiar feature of consciousness: that “there is a constitutive relation between it and knowledge of it” and thus it “can be known independently of any causal impact it makes on a knower, but functional properties cannot” (2009, 295-7). According to Kriegel, this feature of consciousness depends on the fact that “the instantiation of a phenomenal property entails the occurrence of an awareness of that instantiation” (2009, 295), i.e., on conscious form-ness, but the supporter of the QHOT theory may analogously claim that there is a constitutive relation between the property of instantiating mentally quoted intrinsic qualities and the (direct) knowledge of those qualities – which, just like consciousness

for Kriegel, and unlike any other property, can be known independently of their causal impact and their functional profile. Thus, the supporter of the QHOT theory may deny the relevance of the alleged explanatory gap between the relation of cognitive presence and conscious experience – in virtue of which zombies are supposed to still be conceivable – by appealing to the idea that consciousness appears non-functionalizable only because it affords us with a direct cognitive access to intrinsic qualities that can be thus known independently of any functional specification. Accordingly, the supporter of the QHOT theory could maintain that applying Levine’s objection against constitution-based versions of the phenomenal concepts strategy to QHOT theory’s characterization of inner awareness in terms of the physical embedding of first-order states does not lead to the same conclusion because of the less ambitious nature of the QHOT theory: while the physical presence of phenomenal states within phenomenal concepts seemingly cannot fully explain their phenomenal, cognitive presence, the physical presence of first-order states within QHOTs could – thanks to the help of the assumption (unavailable to the supporter of the phenomenal concepts strategy) that the embedded properties of first-order states are already intrinsically qualitative, even when unconscious.

Taking stock, the QHOT theory of consciousness, according to which first-order states acquire a phenomenal character in virtue of being embedded within a quotational higher-order frame that directly displays their qualities to the subject, can allow the higher-order theorist to avoid the illusionist outcomes of representationalist conceptions of inner awareness while preserving the fundamental principles of higher-order intentionalism (i.e., that having consciousness is a matter of inner awareness, and that inner awareness should be conceived in terms of higher-order intentionality) within the framework of the state view. Yet, despite being a perfectly viable position, the QHOT theory will not appear as an attractive view for ‘ambitious’ higher-order theorists who would like higher-order intentionalism to provide sufficient conditions for the constitution of consciousness – by conceiving the intrinsic features of conscious states that give rise to the hard problem in terms of subjective character, rather than in terms of intrinsic qualities. The remainder of this chapter will be devoted to the (speculative) attempt of sketching a higher-order theory that may satisfy this ambition.

## 5.2. The Subject View and a HOP-like Alternative

As considered earlier (§2.1), the formulation of an ambitious higher-order theory involves the characterization of phenomenal character as being made of non-essentially conscious extrinsic properties of mental states that become intrinsic phenomenal qualities in virtue of the instantiation of inner awareness. Such a view can be developed by taking the conscious subjective character of mental states rather than their qualitative character to be primarily responsible for the constitution of their phenomenal qualitativity, i.e., by assuming that “the core of the hard problem is posed not by the qualities themselves but by our experience of these qualities: roughly, the distinctive phenomenal way in which we represent the qualities or are conscious of them” (Chalmers 2018, 30).

The adoption of this explanatory strategy is naturally associated with extrinsic higher-order views – taking consciousness to be a cognitive mechanism that is not part of the phenomenal contents of conscious experiences. For, unless one is ready to embrace the illusionist outcomes of self-representationalism, conceiving subjective character as the intrinsic for-me-ness of phenomenal character naturally leads to abandon higher-order intentionalism – characterizing consciousness as “a sort of *intrinsic glow* that attaches to some mental states and not others” (Kriegel 2009, 101), of the sort defended by (e.g.) Zahavi (1999) and Thomasson (2000). The remainder of this chapter will be devoted to the defense of the (speculative) claim that, instead, within the framework of extrinsic views, it may be possible to explain how an extrinsic property of a first-order state could become an intrinsic phenomenal quality of a conscious state by acquiring an extrinsic kind of subjective character. As considered earlier (§2), since extrinsic state views cannot make room for conscious subjective character without incurring the familiar infinite regress of conscious states, such a view must be developed within the framework of the subject view – according to which mental states become conscious in virtue of being caught up within their subject’s phenomenal perspective.

The proposal presented in what follows will result from the combination of the subject view (§5.2.1) with an attempt to remove the representational layer between the conscious subject and her first-order states from the HOP theory (§5.2.2) – similarly to the way in which the QHOT theory removes that layer from the HOT theory. Clearly, once the representational conception of inner awareness is abandoned, the difference between the



HOP and the HOT theory inevitably narrows down – as it was originally generated precisely by the disagreement on the psychological mode in which the first-order state is represented.<sup>93</sup> For example, explaining consciousness in terms of QHOTs does not require one to posit the problematic extra conditions concerning the aetiology of the relevant higher-order states posited by the HOT theorist: quotational higher-order states are inevitably conceived as assertoric, insofar as they directly “display an *occurrent* [first-order] state to the subject” (Coleman 2015, 2730), for the same reason the non-inferentiality condition is automatically satisfied, and if the QHOT is not simultaneous with the relevant first-order state it will simply not give rise to the conscious experience of that state. Moreover, the contrast between the conceptual nature of thought and the (allegedly) non-conceptual contents of perception – together with its problematic consequences for the HOT theory considered earlier (§4.1.1) – ceases to be relevant, because QHOTs do not determine the features of the experienced contents (Coleman 2015, 2731, fn. 79).

However, there is a significant difference between the HOP and the HOT theory that can survive the removal of the representational layer between the conscious subject and her first-order states: on the HOP theory, but not on the QHOT theory, the existence of inner awareness can be explained by appealing to “the functioning of internal *attention mechanisms* directed upon lower-order psychological states and events” (Lycan 2004, 99). It is this last, remaining difference that will be exploited to develop the present proposal (§5.2.2), after presenting in more detail the idea that first-order states may become conscious in virtue of being caught up within their subject’s phenomenal perspective, rather than in virtue of an unconscious relation with a distinct mental state (§5.2.1.).

### **5.2.1. Higher-order global states**

The idea that consciousness may involve the direct presentation, rather than the representation, of the first-order states made conscious is defended in Van Gulick’s Higher Order Global State (HOGS) model of consciousness (2000; 2004; 2006; 2022). According to the HOGS model, first-order states become conscious in virtue of being

---

<sup>93</sup> In fact, some philosophers even doubt the significance of that distinction within the representationalist framework (e.g., Van Gulick 2000), and Coleman suggests that “it seems possible to read QHOT theory as more HOT- or HOP-like according to one’s preference” (2015, 2730).

suitably integrated into, or recruited by, a global mental state – underpinning the subject’s inner awareness rather than individual fine-grained states. That is, on this view, the higher-order properties responsible for the constitution of inner awareness are conceived as implicit features of the globally integrated state into which first-order states are embedded. Thus, although inner awareness is still conceived in terms of higher-order intentionality, the first-order states made conscious are not taken to be intentional objects of distinct higher-order representations:

The basic idea of the HOGS model is that lower-order object states become conscious by being incorporated as components into the higher-order global states (HOGS) that are the neural and functional substrates of conscious self-awareness. The transformation from unconscious to conscious state is not a matter of merely directing a separate and distinct meta-state onto the lower-order state but of “recruiting” it into the globally integrated state that is the momentary realization of the agent’s shifting transient conscious awareness.

[...] transforming a nonconscious state into a conscious one is a process of recruiting it into a globally integrated complex whose organization and intentional content embodies a heightened degree of reflexive self-awareness. The meta-intentional content is carried not by a distinct and separate vehicle but rather by a complex global state that includes the object state as a component (Van Gulick 2004, 75-7).

Van Gulick specifies the nature of the relevant higher-order content by appealing to phenomenology, arguing that “the phenomenal content of experience extends far beyond what is explicitly present in sensation” and that each phenomenal content involves a high degree of coherence with co-occurrent contents insofar as they all involve the implicit awareness that the conscious subject is a “self located in a world of objects present to it”, and the implicit awareness that the subject’s mental states refer to “a world of objects present from that self’s perspective” (2004, 83).<sup>94</sup> In other words, according to Van Gulick, conscious experiences are essentially framed within a self-world structure that, although generally implicit, is phenomenally conscious nonetheless:

Two interdependent unities pervade the realm of phenomenal experience: the *unity of the experienced world* and the *unity of the experiencing self*. [...] Our phenomenal experience is of a

---

<sup>94</sup> The two are naturally conceived as two sides of the same coin, analogously to inner awareness and cognitive presence.

world of independently existing and relatively stable objects located in a unified arena of space and time within which we ourselves are located as conscious perceivers. [...]

Self and world are two inextricably bound aspects of phenomenal reality. The phenomenal or empirical world is one of objects perceived and experienced from a perspective or point of view, that of the self within that world. Correspondingly the self is always a self set over against a world of objects with which it is engaged. Each side of the phenomenal structure is incoherent without the other (Van Gulick 2004, 81).

That is, on the assumption that we experience reality within a phenomenal structure constituted by the interdependence of a world presented to the subject and the subject to whom that world is presented, it follows that the contents of conscious experiences involve a degree of implicit self-awareness: if it is true that “experience is not a process of blind representation” because the “presence [of the self-world structure] is essential to the content of perceptual experience”, then conscious experience involves a “grasp of that relation” between subject and objects, which in turn implies implicit self-awareness because “understanding the relation requires understanding that to which it is a relation, i.e. understanding it as a relation of presence to self” (Van Gulick 2004, 86).

These phenomenological observations may be supported also by appealing to the embodiment and embeddedness of conscious experience. Plausibly, the properties of a subject’s body determine at least in part which kinds of conscious experiences that subject can undergo, since the body offers at least a causal contribution to many cognitive processes,<sup>95</sup> and the world often serves the same purpose, i.e. the contents of experience vastly depend on the objects we encounter, and many of those objects are often used to offload cognitive effort and enhance our cognitive abilities. These considerations may support Van Gulick’s thesis that phenomenology involves some degree of self-awareness because they point at the fact that we experience the world of objects we perceive as present to us, here and now – and that experience of present-ness might, in turn, essentially involve the awareness of oneself as the entity to whom those present objects are present for.

---

<sup>95</sup> Some philosophers even argue that the body plays a *constitutive* role in cognition, literally as a part of a cognitive system (Shapiro 2019), but this further question has no significant consequence for the present issues.

This idea may be further specified by appealing to the thesis that conscious experiences have *de se* content, i.e., content that conveys non-conceptually to the subject that the subject herself is confronted with certain objects, and that does so without requiring any kind of prior self-identification (Castañeda 1966). The basic intuition behind the notion of *de se* content is that conscious experiences essentially involve two types of ‘aboutness’: one directed towards the experienced (intentional) objects and their properties, and one directed towards the subject of the experience. These two kinds of directedness are supposed to be essentially different in that, in the relevant sense of self-awareness, the subject is not given as another experienced object among others but rather is presented as the very subject of the experience – such that, while I can be mistaken about what is represented as object, it is not clear what it would mean to be wrong about the fact that I am the one undergoing my own experiences. The existence of *de se* content is certainly quite controversial – and providing a full defence of the legitimacy of such a notion is a task that would require a dissertation of its own. However, it should be noticed that an analogous distinction has been drawn in the linguistic domain (Shoemaker 1968), by distinguishing between two ways in which the word ‘I’ can be used: as an object (e.g., ‘I have a broken arm’) and as the subject (e.g., ‘I am in pain’), and pointing out that the latter use of the word ‘I’ is peculiar in that it is immune to error through misidentification. On the one hand, understanding the meaning of sentences in which the word ‘I’ is used as referring to oneself as an object presupposes the recognition of a particular individual – the one with the broken arm – and that task may go wrong in a variety of ways; e.g., I may see my twin with a cast, think I am in front of a mirror, and wrongly conclude that I have a broken arm. On the other hand, it seems that understanding the meaning of sentences in which the word ‘I’ is used as referring to oneself as a subject directly ‘points’ at oneself, without presupposing any prior identification of a presented object as oneself; since, e.g., “to ask ‘are you sure that it’s *you* who have pains?’ would be nonsensical” (Wittgenstein 1958). Analogously, the notion of *de se* content stems from the supposition that there may be a type of mental self-reference that does not require explicit self-identification but is somehow implicitly built into experiences themselves. Similarly, Van Gulick argues that the kind of implicit self-awareness involved in conscious experiences is essentially different from the explicit conscious representation of oneself:

The intentional content of any phenomenal experience always implies the existence of the subject – not merely [blue globe] or even [blue globe here now], but [blue globe seen here/now by me], [blue globe appearing or being present now to me as part of my experienced world]. This is not to say that one is in a state like that one would be in if one explicitly thought or said those words to oneself. The self is not the explicit object of experience in the ordinary case [...] but both the self and its relation to the object are implicit in the structure of the state's phenomenal content. The meta-intentional aspect is built right into the first order content of the experiential state: a dark blue paperweight is present to me as part of my world, i.e. as part of the world that is present from my point of view, which is in turn as self defined by its location in that world of objects and appearance. That sort of implicit reference to self is an essential component of *phenomenal content* [...]. It is part of what distinguishes my *experiencing* the paperweight from merely *representing* it (2004, 85).

Accordingly, on the HOGS model of consciousness the illusionist outcomes of higher-order representationalism are easily avoided, insofar as first-order states are not supposed to become conscious in virtue of being represented – but rather in virtue of being suitably integrated into a global mental state – and the higher-order content allegedly responsible for the constitution of inner awareness is *de se* content (which, unlike ordinary representational contents directed at objects, is immune to error). That is, since on the HOGS model the relevant higher-order content allegedly responsible for the constitution of inner awareness does not need to represent the qualitative richness of experience, there is no risk of misrepresenting it (or representing non-existent features of it). And since the relevant higher-order content is directed at the subject *qua* subject of the experience – i.e., the subject conceived schematically as the entity that is undergoing a certain mental state, rather than as the intentional object of an ordinary representation of oneself – such a higher-order content could not be a misrepresentation of its target (i.e., the subject) without ceasing to give rise to a conscious experience (as it would not represent the subject as the subject of the experience), nor it could lack an existent target (as free-floating experiences without an experiencing subject are arguably metaphysically impossible). Therefore, on the HOGS model of consciousness neither erroneous nor targetless higher-order representations could constitute conscious experiences.

Yet, it is clear that these phenomenological observations can lead to the formulation of a full-fledged theory of consciousness only if the cognitive mechanisms responsible

for the process of integration of first-order states within a global higher-order state is suitably specified, and if such a cognitive mechanism is shown to be sufficient for the constitution of inner awareness. The importance of specifying a cognitive mechanism responsible for the generation of a globally integrated mental state follows from the observation that solely addressing the issue of neurological implementation can hardly provide a satisfying explanation of consciousness:

The global states underlying conscious experience supposedly display a high degree of integration and coherence, but what sorts of coherence are involved and how might they be relevant to the structure of experience and implicit meta-intentionality? Such global states may well cohere in a variety of straightforwardly physical ways [...]. But even if this is so, it [i.e., the relevant neural correlate of the global state] would not by itself provide us with any sort of coherence that helps us with the meta-intentional aspects that we are trying to understand. At most, it may serve as the realization base for some more high-level features that provide the real explanation (2004, 81).

Van Gulick (2006; 2022) starts answering these two questions<sup>96</sup> by appealing to the Global Workspace theory (Baars 1988, 1997; Dehaene 2014), according to which consciousness should be understood in terms of global broadcasting of the information present in individual modular systems in the brain that makes it available to other local modules, thereby allowing them to interact with each other and forming a larger integrated system. Then, he suggests that the global integration of first-order states obtained through the broadcasting of their contents can constitute one's inner awareness of those states by appealing to a broadly functionalist conception of intentional content. That is, on the assumption that a mental state's intentional content depends at least in part on that state's functional role within the cognitive system of which it is part, and given that the global broadcasting of first-order states implies expanding the cognitive system of which they are part, Van Gulick suggests "we should not be surprised if the content of a state shifts when it is embedded within a significantly different context of interactions",

---

<sup>96</sup> That is, the question concerning the nature of the cognitive mechanism responsible for the integration of first-order states, and the question concerning the reason why the presence of such a mechanism should be sufficient for the constitution of inner awareness.

because “if content is a function of functional role, then new contexts that induce states to play new roles may shift their content as well” (2004, 80).<sup>97</sup>

However, developing the HOGS model of consciousness along these lines exposes it to significant objections, casting doubts on whether the global broadcasting of first-order contents can genuinely produce inner awareness. First, as considered earlier while discussing Carruthers’ Dual-Content theory (§4.2), it is likely that the higher-order content acquired in virtue of a first-order state’s mere availability to other local cognitive systems cannot constitute the categorical *feel* of subjective character (at least whenever those higher-order contents are not actually tokened in virtue of the exercise of the subject’s introspective capacities). This objection, however, seems to be less worrisome for Van Gulick than for Carruthers: while on the Dual-Content theory the higher-order content acquired by first-order states (in virtue of their availability to the mind reading system) is supposed to consist in the higher-order representation of that state, which is what allegedly makes up the phenomenal contents of consciousness (and not only their for-me-ness), on the HOGS theory the relevant higher-order content is *de se* content – explaining the subjective givenness of phenomenal states rather than constituting their contents. Accordingly, Van Gulick may simply deny that experiences involve conscious for-me-ness (except for the case of introspection), following the extrinsic higher-order theorist and holding that the subjective givenness of conscious states is not one of their phenomenally manifest features, and that the subject is only unconsciously aware of it – while holding that such implicit awareness of oneself as the subject constitutes inner awareness.

Yet, drawing once again on the previous discussion of Carruthers’ theory, it should be noticed that the possibility of awareness is generally taken to involve actual representation – independently of whether such a representation is conscious or unconscious – and it seems that the sense in which extrinsic higher-order theorists hold that we are (unconsciously) aware of being conscious cannot be genuinely captured by the idea that we know whatever we are disposed to believe (immediately and non-inferentially) in virtue of the global availability of first-order states. However, Van Gulick would likely

---

<sup>97</sup> It should be noticed that the only requirement for defending this thesis is that there is a weak supervenience relation between content and functional role, i.e., that changing the functional role of a mental state can change its content – but that is compatible with the idea that intentional content cannot be reduced to functional role.

agree with this point and reply that, in fact, the dispositional availability of first-order contents only *indirectly* explains their becoming conscious. For, according to him, many unconscious states already possess some degree of implicit higher-order content: “reflexive meta-intentionality is a pervasive and major feature of the mental domain” that “goes way down the phylogenetic scale”, as it is “embedded in procedures that play important intramental roles in producing [...] self-modulation, self-regulation, and self-transformation” (2006, 21).<sup>98</sup> And although “we are able to be conscious in the phenomenal sense at the personal level (or whole organism level) only because we embody such a rich store of implicit and procedural self-understanding at the subpersonal level”, the constitution of “full-blown phenomenal consciousness depends on a *high degree* of implicit self-understanding and meta-intentionality” (2006, 22; emphasis mine) which in turn depends on, but is not to be identified with the global availability of first-order states.

According to Van Gulick, global availability is not consciousness in and of itself, but the former is responsible for the existence of the latter in virtue of the fact that, given the pre-existing implicit higher-order contents of first-order states, “the integrated contents interact with each other and unify with each other *as if from the perspective of a single unified subject, the virtual self*” (Van Gulick 2022, 336). That is, on this view, inner awareness is not directly explained in terms of the (dispositional) availability of first-order states but, rather, is supposed to exist in virtue of the fact that global availability allows the construction of this virtual self – conceived, in a sense, analogously to Dennett’s (1991) self as a virtual centre of narrative gravity, i.e., as “a unified point of view or perspective from which all the contents of experience cohere” – in virtue of which an organism can then become a conscious subject, or a “real self” (Van Gulick 2022, 336). The resulting picture of consciousness resembles the one labelled earlier as the ‘modest subject view’ (§2.1), according to which consciousness is a property of subjects (i.e., on Van Gulick’s view, the property of having a virtual self) that unveils the intrinsic qualities of mental states, thereby transforming them into phenomenal properties, and phenomenal character is constituted by non-essentially conscious intrinsic qualities that

---

<sup>98</sup> For example, he claims that some degree of implicit self-awareness must be involved in the intuitively unconscious cognitive mechanisms responsible for the results of Garcia and Koelling (1967) experiments on rats, showing their learned aversion towards foods eaten in combination with nausea-inducing drugs but not in combination with electric shocks (Van Gulick 2006, 22).



(due to their intrinsic features) can be rendered ‘phenomenal’ in virtue of being caught up within the subject’s phenomenal perspective.<sup>99</sup>

Yet, it is doubtful that explaining the constitution of phenomenal consciousness by appealing to the creation of a virtual self (in virtue of the global broadcasting of already meta-intentional first-order states) can provide sufficient conditions for inner awareness. For, even granting that the presence of such a virtual self could intuitively afford us a conscious phenomenal perspective, it seems that the cognitive mechanism posited by Van Gulick in order to explain its constitution (i.e., global broadcasting) may be active in cases in which the subject is *not* aware of the suitably integrated first-order states. For example, unconscious emotions of a conscious subject are likely to be suitably integrated with co-occurrent conscious states, and yet the subject could lack inner awareness of them:

Consider a situation where I am jealous, but unaware that I’m jealous. I may fancy myself above such petty emotions [...] and even sincerely deny that I’m jealous when asked. But it will be apparent – especially to those close to me – that I am jealous by the way I act. I’ll act rudely towards the person I’m jealous of, I’ll misinterpret his words and actions, I’ll behave aggressively towards him, and so on. Later, I might become conscious of my jealousy, but at the time of the confrontation, I will not be conscious of it. My jealousy in this case certainly appears globally accessible. It controls my moods, my other emotions, my judgements and my perceptions involving the target of my jealousy. It even affects physiological reactions like my temperature and my rate of heartbeat and respiration. The state is not only available to a wide range of systems and processes; it is actively accessed by many of them. Furthermore, I am clearly implicitly self-aware of the state, in Van Gulick’s sense. My jealousy shapes my interactions with my social environment. It determines my judgements, my perceptions, and my behavioural reactions, even my speech. And this in turn feeds back onto my emotional state, affecting its evolution. [...] Therefore, it appears that being a globally accessible state with a high degree of ISA [i.e., implicit self-awareness] is insufficient to make us aware of that state (Weisberg 2008, 178).

Weisberg also suggests further, analogous examples of unconscious states and processes that presuppose global accessibility but do not involve inner awareness, such as the case of unconscious problem-solving processes – in which, depending on the type

---

<sup>99</sup> That is, if realism about the hard problem is presupposed. Otherwise, consciousness could be conceived as the property of subjects responsible for the intrinsic appearance of the properties of first-order states (if illusionism about the hard problem is presupposed), or as the property of subjects responsible for the subject-level accessibility of the properties of first-order states (if eliminativism about the hard problem is presupposed).

of problem at hand, it is likely that “a wide range of systems might be involved in its solution” and that such a solution could not be obtained without “the implicit, harmonizing know-how of ISA”, but “I might come to the solution ‘out of the blue,’ indicating that the problem solving occurred without my being aware of it”; cases of subliminal perception that “can influence patterns of thought and behaviour” as well as “emotional and value judgments”; and cases of chronic pains that despite being sometimes unconscious can permanently “affect mood, focus, intellectual abilities, and so on” (2008, 178). Therefore, it seems likely that a first-order state’s being embedded into a global higher-order state does not guarantee inner awareness of that state, i.e., that HOGS formed by the kind of global broadcasting posited by GWT may have phenomenally unconscious parts. Accordingly, it seems that philosophers sympathetic to the HOGS model of consciousness should seek to individuate alternative cognitive mechanisms able to generate global higher-order states that cannot in principle include unconscious first-order parts. The purpose of the following subsection is to sketch one such proposal, by relying on the basic intuition behind the HOP theory of consciousness, i.e., that “consciousness is the functioning of internal *attention mechanisms* directed upon lower-order psychological states and events” (Lycan 2004, 99), while at the same time rejecting the HOP theory’s traditional conception of inner awareness as constituting representation.

### **5.2.2. *The attention schema: from illusionism to realism***

The basic idea behind my (tentative) proposal is that the integration of first-order states within HOGS, allegedly sufficient for the constitution of inner awareness, may be provided by the ‘attention schema’, i.e., a cognitive model of attention, analogous to the body schema, that allegedly accounts for, or at least improves the top-down control of one’s attention (Graziano 2013; Webb & Graziano 2015; Graziano 2016; 2019). The hypothesis of the existence of an attention schema arises from the intuitive observation that the control of complex systems is greatly improved by the construction of internal simplified models of the mechanisms to be controlled (Camacho & Bordons Alba, 2004). Thus, the idea goes, just like the brain’s control over the body relies on the help of the body-schema – a simplified internal model that contains constantly updating information

about one's body's structure and its current modifications<sup>100</sup> – it is plausible that the brain's control of attention depends at least in part on the presence of a simplified internal model that keeps track of attentional mechanisms and their objects.<sup>101</sup>

Graziano defines the notion of attention as “the ability of the brain to focus its limited resources on a restricted piece of the world at any one time in order to process it in greater depth” – whose basic functioning is conceived in familiar terms as “the ability to enhance some signals over others” (2019, 10) as a result of their competition for ‘cerebral celebrity’ (Dennett 1991) – and the attention schema as “the brain's working description of what it means for a brain to seize on information, focus on it, and deeply process it” (2014, 84), whose alleged functional role is to “monitor the state of attention, keep track of how it can change dynamically from state to state, and predict how it may change in the next few moments (2019, 8). He then distinguishes *overt* attention, dependent on the physical orientation of sensory systems towards the source of salient stimuli, from *covert* attention, intuitively conceived as the “inner spotlight” that “allows us to explore a nearly infinite, multidimensional landscape over which our focus of processing ranges, from the most concrete and immediate objects to the most abstract ideas.” (Graziano 2019, 26), and argues that modelling the cognitive mechanisms responsible for covert attention – given the greater degree of freedom they afford – must involve three fundamental items: not only a representation of the attending subject and a representation of the objects of attention, but also a representation of their relation, “the ever-present process of attention, the computational relationship between the self and everything else” (Graziano 2016, 103), including “the quirky way that attention shifts from place to place, from item to item, its fluctuating intensity, its spatial and temporal dynamics” (Graziano 2013, 83).<sup>102</sup> As mentioned above, these representations are supposed to be simplified models of what is represented, on the assumption that “a detailed, fully accurate internal model is at best wasteful and at worst harmful to the process” (Graziano 2019, 43), i.e., that the power

---

<sup>100</sup> The body schema results from the integration of mainly unconscious proprioceptive information, to be distinguished from a conscious body-image (Gallagher 1998, 228-9), as well as the integration of information provided by the various sense modalities (Graziano and Botvinick, 2002; Armel and Ramachandran, 2003), and plays an important role in the flexible control and short-term planning of action (e.g., Rizzolatti et al., 1996; Graziano and Gross, 1995).

<sup>101</sup> See Graziano (2015, 6-9) for an overview of the empirical evidence in favour of the existence of the attention schema.

<sup>102</sup> The intuition is that once attentional focus becomes dissociated from the physical ‘pointing’ of sensory mechanisms, a model that controls attention cannot only involve the model of the subject (with his sensory mechanisms) and the model of the objects included in the subject's attentional field.

and cognitive efficiency of model-based knowledge precisely depends on the fact that the model does not depict what is modelled in a detailed manner – in the same way in which “the body schema does not depict the mechanistic details that underlie the structure and dynamics of the body” such as “specific bone structure, muscle insertion points, or the molecular basis of muscle contraction” (Graziano 2015, 6).

Graziano develops the hypothesis of the existence of the attention schema into a theory of consciousness by suggesting that consciousness is identical with the attention schema: “subjective awareness *is* the brain’s internal model of the process of attention” (2015, 1; emphasis mine). He suggests that illusionism about the hard problem is the natural outcome of the attention schema theory, in that what we experience is the simplified represented reality contained within such a model, and that this fact “explains why people might mistakenly think that there is a hard problem to begin with, why that mistaken intuition is built deep into us where we’re unlikely to change it, and why its presence is advantageous [...] for the functioning of the brain” (Graziano 2019, 2). However, this illusionist outcome depends on two assumptions: that (i) the contents of the models are representations of salient aspects of first-order states, and (ii) those representational contents constitute the contents of experience.<sup>103</sup> And both assumptions can be rejected – independently of each other – without denying the relevance of the attention schema for the constitution of consciousness, by appealing to the HOGS model of consciousness and conceiving the attention schema as the unifying mechanism responsible for bringing about the integration of first-order states into global mental states.

Assumption (i) may be rejected by claiming that the attention schema, instead of containing representations of selected aspects of first-order states, could embed (and select aspects of) the first-order states themselves. That is, the model of attention could be directly applied ‘over’ those contents, instead of being filled with representations of them, as suggested by the nesting model of mental content considered earlier in this

---

<sup>103</sup> Graziano’s attraction to illusionism also depends on his assumption that crafting a simplified model of the attentional relationship between subject and objects – capturing “the most salient aspects of attention”, such as “the ability to take mental possession of an object, focus one’s resources on it, and, ultimately, act on it”, while overlooking “any of the mechanisms that make this process physically possible” (2015, 2) – implies that the attentional relationship must be represented in such a way that it naturally generates dualist intuitions such that it subjectively seems that there is “an essence that has no specific physical substance but that has a location vaguely inside you” (2019, 42). It is not clear, however, that these two theses are in any particularly tight logical relation. This point will not be pressed further, as what is relevant for present purposes is to establish whether the presence of an attention schema may provide a unifying mechanism for the constitution of HOGS that can lead to develop a theory of consciousness able to capture the intimacy of the relation between subject and her first-order states, instead of only explaining the appearance of intimacy.

chapter. Thus, the attention schema would be conceived as being distributed across the brain (instead of only involving cortical activity as suggested by Graziano 2019), in such a way that the representation of the attentional relationship between subject and objects would involve the presence of a QHOT that directly displays the information concerning the object-relatum (instead of representing it).

Moreover, similar considerations may be also applied to the relevant model of the subject. Graziano suggests that, plausibly, “the body schema and the attention schema [...] partially overlap” (2013, 78), but holds that the model of the subjects involved in the attention schema includes “a model of the self as a physical *and mental* agent” (2015, 2; emphasis mine). However, the characterization of the subject as a mental agent may be only implicitly represented in the attention schema – formed in a way analogous to Van Gulick’s virtual self considered in the previous subsection, in the form of a virtual center of attentional gravity (to borrow Dennett’s phrase); or the subject as a whole (i.e., as a physical *and mental* agent) may be indirectly represented by means of the direct representation of the subject as a physical agent. In both cases, the resulting view would be naturally framed in terms of the ambitious subject view, as consciousness could not be simply reduced to the relationship between QHOT and quoted first-order states made conscious; rather, it would be primarily constituted by the property of the subject of having a cognitive structure (i.e., the attention schema) able to integrate one’s first-order states into HOGS in virtue of the representation of one’s attentional relationship with those states. That is, consciousness would consist of two distinct kinds of phenomenal properties: a phenomenal perspective, conceived as the property of the subject of having an attention schema made conscious by the simplified representation of the attentional relationship between subject and (intentional) objects, and the phenomenal character of the quoted first-order states – unified in virtue of being caught up within the subject’s phenomenal perspective. Thus, for example, the fact that there is something it is like to see the blue sky would be explained by the fact that the first-order representation of the blue sky enters the attentional field of the subject (by being quoted by the attention schema) *and* by the fact that the subject has a phenomenal perspective on it (granted by the higher-order representation of the subject’s attentional relationship with it).

Alternatively, it may be possible to accept that the contents of the attention schema are in fact representations of first-order states while rejecting assumption (ii), i.e., denying

that the phenomenal contents of experience should be identified with the representational contents of the attention schema. That is, the simplified representations involved in the construction of the attention schema may be instrumental to the control of top-down attention even if those representations remained mainly unconscious – analogously to the case of most proprioceptive information present in the body schema which, nevertheless, are supposed to help control one’s body (Gallagher 1998, 228-9).<sup>104</sup> The resulting view would be, once again, naturally conceived as an ambitious subject view. Consciousness could not be simply reduced to a relationship between the attention schema and first-order states, because its constitution would depend on the cognitive integration brought about by the presence of such a relation (rather than directly on the relation itself). On this view, consciousness would consist in the subject’s capacity of consciously relating to her first-order states in virtue of the property of having an unconscious attention schema – which would endow the subject with HOGS by producing the subject-object structure into which the contents competing for cerebral celebrity are embedded (rather than generating the contents of conscious experience as suggested by Graziano). For example, the fact that there is something it is like to see the blue sky would be explained by the fact that the subject unconsciously represents a (higher-order) simplified version of the represented blue sky as being related to oneself through one’s attention, in virtue of which the subject allegedly becomes able to consciously relate the (original) visual representation of the blue sky as the subject of that representation.

That is, the unconscious knowledge of being the subject of experience – i.e., of being a self that is related to the intentional objects of experience by means of one’s attentional mechanisms – could make the subject act in a conscious way: taking a subjective (mental) stance towards the contents of one’s mental states, which in turn could constitute those content’s being conscious (i.e., inner awareness). Thus, even though the attention schema itself would be unconscious, the exercise of this capacity could constitute a conscious phenomenal perspective, because the implicit higher-order content acquired by those occurrent first-order states in virtue of their global integration would not be simply a dispositional matter as in Van Gulick’s proposal but, rather, it would depend on the

---

<sup>104</sup> In a sense, the resulting view would be reminiscent of Block’s (2011b) view that phenomenal consciousness “overflows” cognitive access, on the assumption that accessibility depends on the presence of the attention schema.

presence of an explicit (albeit unconscious) representation of oneself as being related to one's mental states.

As a result, the unconscious attention schema could make a phenomenal contribution to conscious experiences by imposing a unifying structure on the occurrent first-order states that can become objects of attentional focus (i.e., the mental states encompassed within the subject's attentional field, competing for cerebral celebrity). However, such a phenomenal contribution would still be only implicit in the phenomenal character of the mental states made conscious: it would consist in the subject's awareness of being in an awareness relation with those states, rather than being an aspect of their phenomenal contents, i.e., it would amount to a conscious phenomenal perspective, rather than to an explicit feature of the phenomenal character of the states caught up within such a structure.

Therefore, on either of these possible non-illusionist uses of the notion of attention schema in order to explain the integration of first-order states into HOGS and the constitution of inner awareness, consciousness would be conceived as being primarily a property of subjects (insofar as it could not be reduced to a relation between subpersonal states), and it might be explained without positing the existence of intrinsic properties of unconscious mental states but, rather, by taking the extrinsic subjective character of conscious states to be responsible for the constitution of their phenomenal qualitativity – on the assumption that “the core of the hard problem is posed not by the qualities themselves but by our experience of these qualities: roughly, the distinctive phenomenal way in which we represent the qualities or are conscious of them” (Chalmers 2018, 30). Clearly, one may reject this assumption and hold that, for independent reasons (e.g., the possibility of inverted spectra), intrinsic qualities must be posited nonetheless. However, the proposals under consideration would still appear at least preferable to Van Gulick's articulation of the HOGS theory, since they provide a unifying mechanism – affording the subject with the cognitive context required for inner awareness (i.e., the subject-object structure described earlier) – superior to the virtual self allegedly constituted by the global

broadcasting of first-order contents, in that such mechanisms do not allow for suitably integrated and yet unconscious first-order states, as Van Gulick's HOGS do.<sup>105</sup>

Moreover, the appeal of these alternative conceptions of the consciousness-constituting HOGS may be increased on the (phenomenologically based) assumption that phenomenal consciousness involves conscious subjective character, or the feel of awareness (i.e., conscious awareness of awareness). Granted, the present conception of (extrinsic) subjective character as conscious phenomenal perspective, rather than intrinsic for-me-ness of conscious states, is rather unorthodox. Yet, it can arguably capture the intrinsic theorists' descriptions of phenomenology, and also offer a significant theoretical advantage, in that it provides a more plausible characterization of what conscious inner awareness could consist in.

The possibility of intrinsic for-me-ness is questioned by Coleman (2017) on the grounds that it would involve an implausible phenomenal "duplication" of qualities: "the alleged feel of awareness" is construed as an "additional sensory content beyond the other qualities one is aware of", i.e., as another item in phenomenology (e.g., Kriegel 2009), it is natural for the intrinsic theorist to suppose that such a feel is somehow "suffused with the qualities that awareness is of" (2017, 272). But then it seems to follow that "I get every first-order quality twice in consciousness: once in its own right (as a 'floor-level item,' in Kriegel's phrase), and once more as 'staining' the feel of my awareness of all these first-order qualities", insofar as "the sensory quality of awareness is posited as an item additional to the first-order qualities, while containing, in its feel (where else?), reference to them" – despite the implausibility that this "doubling of qualities" is actually displayed in phenomenology (Coleman 2017, 272-3).

Arguably, the intrinsic theorist cannot avoid this consequence, since once it is assumed that subjective character is an intrinsic feature of conscious states' phenomenal character, it naturally follows that the former is given as an aspect of the qualities determining the identity of the latter (e.g., Kriegel 2009, 11) That is, if subjective character had "its own, 'isolated,' feel", making it "a standalone qualitative ingredient in consciousness" rather

---

<sup>105</sup> For example, one's unconscious jealousy, despite its influence on various local systems and processes, would not be considered as being part of the consciousness-supporting HOGS as it is out of the scope of one's attentional field. Analogous considerations apply to the other examples (unconscious problem-solving, subliminal perception, and unconscious chronic pains) considered earlier in Weisberg's (2008) objection against Van Gulick's articulation of the HOGS model.



than a peculiar feel characterized by its being “somehow interpenetrated by the other, first-order, qualities of which one is aware”, it would be “very hard to see how, in experiencing this quality, one could apprehend it as a feeling of awareness of these (first-order) qualities, i.e., as the very item it is supposed to be” (Coleman 2017, 272).

Therefore, on intrinsic higher-order theories, it seems that awareness of awareness cannot figure in phenomenology as including the qualities it is about – on pain of implausible quality duplication – nor as a general awareness-feel – as it would make “the feel of awareness of first-order qualities unidentifiable as such, and likely wholly mysterious: a detached phenomenal UFO” (Coleman 2017, 272). By contrast, it seems that on the proposals under consideration it becomes possible to conceive the phenomenal contribution of subjective character as a general awareness-feel while still recognizing it as the feel of one’s awareness of first-order qualities (i.e., without making the feel of awareness unidentifiable nor mysterious), as it is not conceived as another item included in the phenomenal character of the mental state made conscious but, rather, as the feel of *being conscious*, of *having* inner awareness. The present suggestion is that the experience of (extrinsic) subjective character is primarily the experience of having a conscious phenomenal perspective, rather than the intrinsic for-me-ness of conscious states. But such an experience could still be described as the experience of one’s awareness of first-order qualities, insofar as the feel of being an experiencing subject and the cognitive presence of the experienced mental qualities may be seen as two sides of the same coin. That is, on this view, the feel of awareness described by Kriegel (2009) as elusive and only peripherally the object of awareness would be the feel of being an experiencing subject – not something intrinsic to the phenomenal character of the mental states made conscious – but it would also be, *implicitly*, the feel of our awareness of qualities, which could in turn be made explicit in introspection. Thus, when introspecting subjective character, it would appear as if it were an intrinsic feature of phenomenal character (without giving rise to phenomenal duplication, as the qualities would only be experienced *through* their for-me-ness one is introspecting), as generally described by intrinsic theorists, but one’s ordinary experience of it would not be suffused with qualities insofar as it would be characterized as the other side of the same coin (i.e., consciousness of having inner awareness, instead of consciousness of inner awareness of the qualities one is aware of). Clearly, these suggestions are highly speculative and would require a

much more thorough defense. However, it seems that they may lead to a promising explanatory strategy – able to account for the alleged feel of awareness, and thus for the distinctive way in which mental states’ properties are given in conscious experience.

In conclusion, it may be asked whether such an unorthodox variation of the higher-order approach should be regarded as a genuinely meta-intentional theory of consciousness – a question sometimes asked about Van Gulick’s HOGS theory as well. For, differently from traditional higher-order representational theories as well as from self-representationalism and the QHOT theory, on these views inner awareness does not wholly consist in the instantiation of higher-order intentionality.<sup>106</sup> In the case of the HOGS theory, the extra ingredient is the cognitive integration of unconscious first-order and higher-order contents into a global mental state, by means of which the subject allegedly acquires a ‘virtual self’ that constitutes his phenomenal perspective. In the case of the alternative developments of the HOGS model just presented, the extra ingredient is the cognitive integration of unconscious first-order contents (or of simplified representations of those contents) within an attention schema, by means of which the subject allegedly acquires awareness of his attentional relationship with the first-order contents thereby made conscious (or acquires the capacity of consciously relating to those contents). However, even though both versions of the HOGS model rely on non-intentional notions and thus imply that consciousness is not identical with meta-intentionality, they still remain faithful to the spirit of higher-order intentionalism: explaining consciousness in terms of inner awareness – as suggested by the transitivity principle – and grounding the existence of inner awareness on meta-intentionality. According to the HOGS theory, if unconscious mental states did not already possess *de se* content – concerning the subject *as the subject* of that state, and responsible for the constitution of the virtual self when suitably integrated – the existence of consciousness would be impossible. Analogously, according to the alternative developments of the HOGS model just presented, without the subject-object structure produced by the attention schema – into which the first-order contents competing for cerebral celebrity are embedded – and without the (higher-order) representation of the attentional relationship between the two, consciousness would not arise. Therefore, in both cases, although

---

<sup>106</sup> Thanks to Tom McClelland for raising this point.

consciousness is not taken to be reducible to higher-order intentionality, the existence of consciousness is certainly grounded on, and explained by intentionality. And, ultimately, that is all that matters to take a step further in the quest for a naturalistic understanding of consciousness.

## Conclusions

The purpose of this dissertation was to provide an analysis of the possible approaches to the hard problem of consciousness within the framework of higher-order intentionalism – according to which phenomenal consciousness consists in the subject’s inner awareness of her mental states – and to assess their prospects of delivering satisfying accounts of conscious experience. Two main kinds of higher-order theories have been distinguished:

According to extrinsic higher-order theories, traditionally exemplified by Rosenthal’s (1986) HOT theory and by Armstrong’s (1980) and Lycan’s (1987) HOP theory, inner awareness is distinct from the experienced properties of the mental state made conscious and constitutes our experience of them.

According to intrinsic higher-order theories, notably exemplified by Kriegel’s (2009) Self-Representational theory, inner awareness is an intrinsic feature of phenomenal character, constituted by the experienced properties of the mental state that becomes conscious.

The distinction between these two kinds of higher-order theory has been further spelt out by considering the two questions any theory of consciousness must answer:

- (a) What kind of properties constitute the contents of conscious experience?
- (b) What kind of properties make a subject conscious?

While the intrinsic theorist proposes to answer question (b) derivatively, in terms of one’s answer to question (a), the extrinsic theorist holds that one’s answer to question (b) is somewhat explanatory prior to one’s answer to question (a) for the purpose of explaining the existence of inner awareness – either because question (a) can be answered in terms of one’s answer to question (b), as suggested by ambitious extrinsic theories, or because answering question (a) does not provide *ipso facto* an answer to question (b), as suggested by modest extrinsic theories. That is, according to the intrinsic theorist, devising a theory of consciousness means explaining how mental states acquire essentially conscious properties (i.e., qualities-for-me-ness, in the case of higher-order intrinsic theories) that establish a relation of inner awareness with the subject instantiating them. By contrast, according to the extrinsic theorist, devising a theory of consciousness

means explaining the nature of the cognitive mechanism responsible for the constitution of inner awareness, and then the subject's inner awareness is taken to make one's mental states conscious – either by unveiling first-order states' qualities (in the case of modest views) or by constituting their phenomenal qualitativity (in the case of ambitious views).

In assessing the prospects of these two kinds of higher-order theories, I have argued that the framework of extrinsic theories allows the higher-order theorist to adopt a wider variety of explanatory strategies to tackle the hard problem, unavailable within the framework of intrinsic theories.<sup>107</sup>

Intrinsic higher-order theories are naturally interpreted as being committed to the state view, according to which consciousness is primarily a property of mental states: since the properties making the subject conscious are supposed to be intrinsic to the phenomenal character of the conscious state, the property of the subject of being phenomenally presented with certain experiential qualities (and their alleged for-me-ness) is characterized as derivative upon the property of having mental states with essentially conscious qualities making the subject conscious of themselves. In turn, this feature of intrinsic higher-order theories may appear appealing to some philosophers, because it suggests that those theories promise to provide a plausible realist solution to the hard problem – according to which consciousness involves the instantiation of intrinsic phenomenal properties. However, it is doubtful that they can succeed. For, unless the intrinsic theorist is ready to abandon the core tenet of higher-order theories (i.e., that consciousness should be explained in terms of higher-order intentionality), it seems that the formulation of an intrinsic theory presupposes one's commitment to a (constituting) representationalist conception of inner awareness (Kriegel 2009, 99-113). But, I have argued that, in turn, intrinsic higher-order representationalism – according to which conscious states are supposed to be complexes, formed by the sum of first-order states made conscious and their higher-order representations – naturally leads to develop a theory of consciousness that is committed to an illusionist approach to the hard problem – according to which consciousness only involves the (extrinsically constituted) appearance of intrinsic phenomenal properties of the mental states made conscious. For the intrinsic phenomenal qualities allegedly attributed to first-order states (in virtue of

---

<sup>107</sup> As well as a wider variety of metaphysical positions concerning the fundamental nature of properties.

their being part of complex conscious states) turn out to be irrelevant to the constitution of conscious experience: on intrinsic higher-order representationalism, phenomenal consciousness is only constituted by the subjective impression of having conscious first-order states – determined by the presence of higher-order representations that indirectly represent themselves by directly representing possibly non-existent first-order contents.

By contrast, extrinsic theories have been traditionally articulated within the framework of the state view (explaining state-consciousness in terms of the representational relation between two subpersonal mental states) and as involving an eliminativist attitude towards the hard problem – according to which consciousness does not involve the instantiation of intrinsic phenomenal properties by the mental states made conscious (nor the appearance that such properties are instantiated). Yet, the explanatory strategy characteristic of extrinsic higher-order theories is compatible with illusionist and realist approaches to the hard problem, as well as with the subject view – according to which consciousness is primarily a property of subjects and only derivatively a property of mental states, such that the property of having mental states with phenomenal qualities is characterized as derivative upon the property of the subject of being phenomenally presented with some of her mental states' properties.

On the one hand, extrinsic higher-order representationalism can be interpreted as a form of illusionism by adopting the subject view and holding that the higher-order representations of first-order states (in virtue of which the subject allegedly becomes aware of being in those states) do not make, strictly speaking, their intentional objects (state-)conscious but, rather, they are themselves the conscious states (constituting a fundamental kind of creature-consciousness). That is, rather than conceiving consciousness as the property of first-order states of becoming conscious in virtue of being intentional objects of higher-order unconscious representations, the extrinsic theorist may hold that first-order states appear as having intrinsic for-me-ness in virtue of the property of the subject of representing her own mental life (e.g., Brown 2015).<sup>108</sup>

---

<sup>108</sup> In this way, the extrinsic theorist may also avoid the intrinsic theorist's commitment to the thesis that consciousness is a categorical property, by characterizing consciousness as the power of representing possibly nonexistent aspects of one's mental life. By contrast, the intrinsic higher-order representationalist would still be committed to the claim that those represented first-order contents only become conscious in virtue of the categorical (*qua* aspect of phenomenal character) property of the relevant higher-order representation of indirectly representing itself.

On the other hand, extrinsic higher-order theories can be made compatible with realist attitudes towards the hard problem by renouncing the constituting representationalist conception of inner awareness without giving up higher-order intentionalism.

Such a task may be carried out within the framework of the modest extrinsic view – according to which consciousness should be conceived as the property of first-order states of being object of unconscious higher-order intentional states that unveil their pre-existing intrinsic qualities, or as the property of the subject of being phenomenally presented with those qualities in virtue of having unconscious higher-order states about them – by conceiving the consciousness-generating higher-order states as non-representational, quotational mechanisms that directly exhibit first-order states to the subject, by physically embedding them within frame-like structures (Coleman 2015).<sup>109</sup>

Alternatively, it may be possible to develop ambitious extrinsic higher-order theories – according to which consciousness is the property of subjects of having a phenomenal perspective that constitutes the phenomenal qualitativity of the first-order states caught up within it – by appealing to Van Gulick’s HOGS model of consciousness – according to which first-order states become conscious in virtue of being suitably integrated into, or recruited by, a global mental state. In particular, I have suggested that the integration of first-order states should not be explained in terms of the Global Workspace Theory (Baars 1997) as suggested by Van Gulick and that, instead, it may be due to the presence of an attention schema, i.e., a cognitive model of attention supposed to improve the top-down control of it (Graziano 2013) – which may constitute conscious inner awareness either by combining the (simplified) representation of the subject’s attentional relationship with her first-order states and the higher-order quotation of their contents, or by producing an unconscious subject-object structure, into which the contents competing for cerebral celebrity are embedded, that may allow the subject to take a subjective stance towards those contents (thereby generating conscious inner awareness of them).

The main appeal of these sketched proposals consists in their ability to provide plausible characterizations of conscious inner awareness – whose existence must be denied by the supporter of extrinsic state views (due to the familiar infinite regress of

---

<sup>109</sup> In this way, the extrinsic theorist may avoid the intrinsic theorist’s commitment to the thesis that consciousness is a categorical property by characterizing consciousness as the power of subjects of unveiling the intrinsic qualitative aspects of mental states (thereby turning them into phenomenal properties).

conscious states) – alternative to the problematic intrinsic for-me-ness posited by intrinsic higher-order theories. For, on these views, it becomes possible to conceive the subject as being conscious of one’s awareness as the peculiar way in which one is related to one’s own mental state through consciousness, rather than as a specific phenomenal content ascribed to the mental states made conscious (e.g., Nida-Rümelin 2014). That is, subjective character could be, indeed, an intrinsic aspect of every conscious experience, but without being part of its content: it would amount to what it is like for the subject to perform the cognitive activity that gives her conscious access to some of her mental states (instead of what it is like for the subject to be aware of the objects of that cognitive activity, i.e., first-order contents). In this way, phenomenology could be characterized as involving conscious features that are irreducible to the property of the mental states made conscious of having a phenomenal character, by appealing to the idea that consciousness may be conceived as the cognitive structure connecting the heterogeneous properties of mental states into a unified conscious perspective on one’s mental life – such that one’s conscious experiences would not be entirely constituted by properties of the interconnected parts of one’s mental life, but rather by the property of subjects bringing them together (i.e., a subject’s phenomenal perspective).<sup>110</sup>

Therefore, given the wide variety of plausible approaches to consciousness available to the extrinsic theorist but unavailable to the intrinsic theorist, it seems that the explanatory strategy associated with extrinsic higher-order theories ultimately provides the most fruitful framework for philosophers wanting to explain phenomenal consciousness while assuming the truth of higher-order intentionalism.

---

<sup>110</sup> In this way, the extrinsic theorist may also avoid the intrinsic theorist’s commitment to the thesis that consciousness is a categorical property by characterizing consciousness as the power of subjects of constituting the subjective character of conscious experiences.



## References

- Alter T, (2016) The Structure and Dynamics Argument against Materialism. *Nouûs* 50: 794-815.
- Armel KS, Ramachandran VS, (2003) Projecting sensations to external objects: evidence from skin conductance response. *Proc Biol Sci* 270: 1499-1506.
- Armstrong D, (1980) What is Consciousness? In *The Nature of Mind*, Armstrong, D, pp 55-67. Australia: University of Queensland Press.
- Armstrong D, (1997) *A World of States of Affairs*. New York: Cambridge University Press.
- Baars B, (1988) *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Baars B, (1997) *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.
- Balog K, (2012) Acquaintance and the mind-body problem. In *New Perspectives on Type Identity: The mental and the physical*, Hill C, Gozzano S, (eds) pp 16-43. Cambridge: Cambridge University Press.
- Bayne T, Chalmers D, (2003) What is the unity of consciousness? In *The Unity of Consciousness: Binding, Integration, and Dissociation*, Cleeremans A, (ed) pp 23-58. Oxford: Oxford University Press.
- Bayne T, (2007), Conscious States and Conscious Creatures: Explanation in the Scientific Study of Consciousness. *Philosophical Perspectives* 21: 1-22.
- Berto F, Schoonen T, (2018) Conceivability and Possibility: Some Dilemmas for Humeans. *Synthese* 195: 2697–2715.
- Bird A, (1998) Dispositions and Antidotes. *Philosophical Quarterly* 48: 227-234.
- Bird A, (2007) The Regress of Pure Powers? *Philosophical Quarterly* 229: 513-534.
- Blackburn S, (1990) Filling in Space. *Analysis* 50: 62-5.

Block N, (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227-287.

Block N, (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 30: 481-548.

Block N, (2011a) The higher order approach to consciousness is defunct. *Analysis* 71: 419-431.

Block N, (2011b) Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* 15: 567-575.

Block N, (2015) The Puzzle of Perceptual Precision. In *Open MIND*, Metzinger T, Windt JM (Eds) pp 167-218. Frankfurt am Main: MIND Group.

Byrne A, (1997) Some like It Hot: Consciousness and Higher-Order Thoughts. *Philosophical Studies* 86: 103-129.

Brown R, (2015) The HOROR Theory of Phenomenal Consciousness. *Philosophical Studies* 172: 1783-1794.

Camacho EF, Bordons Alba C, (2004) *Model Predictive Control*. New York: Springer.

Carruthers P, (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Carruthers P, (2005) *Consciousness: Essays From a Higher-Order Perspective*. Oxford: Oxford University Press.

Carruthers P, (2019) *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford: Oxford University Press.

Carruthers P, Gennaro RJ, (2020) Higher Order Theories of Consciousness. *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/fall2020/entries/consciousness-higher/>.

Castañeda H, (1966) 'He': A Study in the Logic of Self-Consciousness. *Ratio* 8: 130–57.

Caston V, (2002) Aristotle on Consciousness. *Mind*, 111: 751-815.

Chalmers D, (1996) *The Conscious Mind*, New York: Oxford University Press.

Chalmers D, (2004) The Representational Character of Experience. In *The Future for Philosophy*, Leiter B, (ed) pp 153-181. New York: Oxford University Press.

Chalmers D, (2016) Panpsychism and panprotopsyism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, Alter T, Nagasawa Y, (eds) pp 246-276. New York: Oxford University Press.

Chalmers D, (2018) The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, 25: 6–61.

Churchland PS, (1983) Consciousness: The Transmutation of a Concept. *Pacific Philosophical Quarterly* 64: 80-95.

Coates P, Coleman S, (2015) The Nature of Phenomenal Qualities. In *Phenomenal Qualities*, Coates P, Coleman S (eds) pp 1-33. New York: Oxford University Press.

Coleman S, (2014) The real combination problem: Panpsychism, micro-subjects, and emergence. *Erkenntnis* 79: 19-44.

Coleman S, (2015) Quotational higher-order thought theory. *Philosophical Studies* 172: 2705-2733.

Coleman S, (2017) Panpsychism and Neutral Monism. In *Panpsychism: Contemporary Perspectives*, Bruntrup G, Jaskolla L (eds) pp 249-282. New York: Oxford University Press.

Coleman S, (2018) The merits of higher-order thought theories. *Transformação Revista de Filosofia* 41: 31-48.

Coleman S, (2019) Natural Acquaintance. In *Acquaintance: New Essays*, Knowles J, Raleigh T (eds) pp 49-74. New York: Oxford University Press.

Coleman S, (2022) Intentionality, Qualia, and the Stream of Unconsciousness. *Phenomenology and Mind* 22: 42-53.

Crane T, (2001) *Elements of mind*. New York: Oxford University Press.

Crick F, & Koch C, (1990) Towards a Neurobiological Theory of Consciousness. *Seminars in Neuroscience 2*: 263-75.

Dainton B, (2008) *The Phenomenal Self*. Oxford: Oxford University Press.

Damasio A, (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Boston, MA: Mariner Books.

Davidson D, (1969) The individuation of events. Reprinted in *Essays on Actions and Events*, Davidson D, (2001) pp 163-180. Oxford: Oxford University Press.

Davidson, D. (1970). Events as particulars. Reprinted in *Essays on Actions and Events*, Davidson D, (2001) pp 181-188. Oxford: Oxford University Press.

Dehaene S, (2014) *Consciousness in the brain*. New York: Penguin.

Dennett D, (1978) *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. (1991). *Consciousness Explained*. New York: Little, Brown & Co.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Ellis B, (2001) *Scientific Essentialism*, Cambridge: Cambridge University Press.

Feinberg TE, (2000) The nested hierarchy of consciousness: A neurobiological solution to the problem of mental unity. *Neurocase 6*: 75–81.

Feinberg TE, (2005) Neural Hierarchies and the Self. In *The Lost Self: Pathologies of the Brain and Identity*, Feinberg TE, Keenan JP, (eds) pp 33-49. New York: Oxford University Press.

Fine K, (1994) Essence and Modality. *Philosophical Perspectives 8*: 1-16.

Frankish K, (2016) Illusionism as a theory of consciousness. *Journal of Consciousness Studies 23*: 11–39, Reprinted in *Illusionism as a Theory of Consciousness*, Frankish K (2017) pp 13-48, Imprint Academic.

Gallagher S, (1998) Body schema and intentionality. In *The Body and The Self*, Bermudez JL, Marcel A, Glan N, (eds) pp 225–244. Cambridge, MA: MIT Press.

Garcia JA, Koelling RA, (1967) A comparison of aversions induced by X rays, toxins, and drugs in the rat. *Radiation Research Supplement 7*: 439-450.

Gennaro RJ, (1996) *Consciousness and Self-Consciousness*. Amsterdam: John Benjamins.

Gennaro RJ, (2012) *The Consciousness Paradox: consciousness, concepts, and High-Order Thoughts*. Cambridge, MA: MIT Press.

Giannotti J, (2019). The Identity Theory of Powers Revised. *Erkenntnis* 86: 603-621.

Goldman A, (1993) Consciousness, folk-psychology, and cognitive science. *Consciousness and Cognition* 2: 364-382.

Graziano MSA, Gross CG, (1995) Multiple representations of space in the brain. *Neuroscientist* 1: 43-50.

Graziano, MSA, Botvinick MM, (2002) How the brain represents the body: Insights from neurophysiology and psychology. In *Common Mechanisms in Perception and Action: Attention and Performance*, Prinz J, Hommel B, (eds) pp. 136–157. Oxford: Oxford University Press.

Graziano MSA, (2013) *Consciousness and the Social Brain*. New York: Oxford University Press.

Graziano MSA, (2016) Consciousness engineered, *Journal of Consciousness Studies*, 23: 98-115.

Graziano, MSA. (2019). *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. New York: W.W. Norton & Co.

Heil J, (2003) *From an Ontological Point of View*. New York: Oxford University Press.

Heil J, (2010) Powerful qualities. In *The metaphysics of powers: Their grounding and their manifestations*, Marmodoro A, (ed) pp 58-72. New York: Routledge.

Heil J, (2012) *The Universe as We Find it*. New York: Oxford University Press.

Hill C, (2009) *Consciousness*. Cambridge: Cambridge University Press.

Horgan T, (1993) From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind* 102: 555-586.

Howell R, (2008) The Two-Dimensionalist Reductio. *Pacific Philosophical Quarterly* 89: 348-358.

Humphrey N, (2016) Redder than red: Illusionism or phenomenal surrealism? *Journal of Consciousness Studies* 23: 116-123.

Hurley S, (1998) *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Ingthorsson RD, (2012) The Regress of Pure Powers Revisited. *European Journal of Philosophy* 23: 529-541.

Irvine L, Sprevak M, (2020) Eliminativism about consciousness. In *Oxford Handbook of the Philosophy of Consciousness*, Kriegel U (ed) pp 348-370. Oxford: Oxford University Press.

Jackson F, (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Johnston M, (1992) How to Speak of the Colors. *Philosophical Studies* 68: 221-263.

Kim J, (1973) Causation, nomic subsumption and the concept of event. Reprinted in *Supervenience and mind*, Kim J, (1993) pp 3-21. Cambridge: Cambridge University Press.

Kim J, (1976) Events as property exemplifications. Reprinted in *Supervenience and mind*, Kim J, (1993) pp 33-52. Cambridge: Cambridge University Press.

Kim J, (1982) Psychophysical Supervenience. *Philosophical Studies* 41: 51-70.

Kim J, (1993) *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kind A, (2003) What's So Transparent About Transparency? *Philosophical Studies* 115: 225-244.

Kriegel U, (2009) *Subjective Consciousness: A Self-Representational Theory*. New York: Oxford University Press.

Kriegel U, (2012) In defense of self-representationalism: reply to critics. *Philosophical Studies* 159: 475-484.

Leslie AM, (1987) Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review* 94: 412-426.

Levine J, (1983) Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* 64: 354-61.

Levine J, (2001) *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

Levine J, (2006) Phenomenal Concepts and the Materialist Constraint. In *Phenomenal Concepts and Phenomenal Knowledge: New essays on consciousness and physicalism*, Alter T, Walter S, (eds) pp 145-166. Oxford: Oxford University Press.

Lewis D, (1983) Extrinsic Properties. *Philosophical Studies* 44: 197-200.

Lewis D, (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lycan W, (1987) *Consciousness*. Cambridge, MA: MIT Press.

Lycan W, (1996) Consciousness as Internal Monitoring. In *The Nature of Consciousness*, Block N, Flanagan O, Guzeldere G, (eds) pp 755-772. Cambridge, MA: MIT Press.

Lycan W, (2004) The superiority of HOP to HOT. In *Higher-Order Theories of Consciousness*, Gennaro, R (ed) pp 93-114. Philadelphia: John Benjamins.

Lycan W, (2008) Phenomenal Intentionalities. *American Philosophical Quarterly* 45: 233-252.

Lockwood M, (1989) *Mind, brain and the quantum: The Compound 'I'*. Oxford: Basil Blackwell.

Lowe EJ, (2006). *The four category ontology*. Oxford: Oxford University Press.

Mayr E, Marmodoro A, (2019) *Metaphysics*. New York: Oxford University Press.

Marmodoro A, (2009) Do Powers Need Powers to Make Them Powerful? From Pandispositionalism to Aristotle. *History of Philosophy Quarterly* 26: 337-352.

Marmodoro A, (2010) *The Metaphysics of Powers: Their Grounding and their Manifestations* (ed). New York: Routledge.

Marmodoro A, (2017) Aristotelian Powers at Work: Reciprocity without Symmetry in Causation. In *Causal Powers*, Jacobs J (ed) pp 57-76. Oxford: Oxford University Press.

Marmodoro A, (2020) Powers, Activity and Interaction. In *Dispositionalism: Perspectives from Metaphysics and the Philosophy of Science*, Meincke AS (ed) pp 55-66. Cham: Springer.

Martin CB, (1993) Power for realists. In *Ontology, Cause and Mind: Essays in honour of D. M. Armstrong*, Bacon J, Campbell K, Reinhardt L, (ed) pp 175-786. Cambridge: Cambridge University Press.

Martin CB, (1994) Dispositions and Conditionals. *Philosophical Quarterly* 44: 1-8.

Martin CB, (1996) Properties and dispositions. In *Dispositions: A debate*, Crane T, (ed) pp 71-87. New York: Routledge.

Martin CB, (1997) On the Need for Properties: The Road to Pythagoreanism and Back. *Synthese* 112: 193-231.

Martin CB, Heil J, (1999) The ontological turn. *Midwest Studies in Philosophy*, 23: 34–60.

McDowell J, (1994) *Mind and World*. Cambridge MA: Harvard University Press.

Mill JS, (1967) *A System of Logic*. London: Longmans.

Mihálik J, (2022) Panqualityism, Awareness and the Explanatory Gap. *Erkenntnis* 87:1423–1445.

Molnar G, (2003) *Powers*. New York: Oxford University Press.

Mumford S, Anjum R, (2011) *Getting Causes from Powers*. New York: Oxford University Press.

Nagel T, (1974) What Is It Like to Be a Bat? *The Philosophical Review* 83: 435-450.



Neander K, (1998) The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness. *Philosophical Perspectives* 12: 411-434.

Nida-Rümelin M, (2014) Basic Intentionality, Primitive Awareness, and Awareness of Oneself. In *Mind, Values, and Metaphysics: Philosophical Essays in Honor of Kevin Mulligan Vol.2*, Reboul A (ed) pp. 261-290. Switzerland: Springer International.

Nida-Rümelin M, (2017) Self-awareness. *Review of Philosophy and Psychology* 8: 55–82.

Nida-Rümelin M, (2018) The experience property framework: A misleading paradigm. *Synthese*, 195: 3361–3387.

O'Brien G, Opie J, (1998) The disunity of consciousness. *Australasian Journal of Philosophy* 76: 378–395.

Papineau D, (2002) *Thinking about consciousness*. Oxford: Oxford University Press.

Papineau D, (2006) Phenomenal and perceptual concepts. In *Phenomenal Concepts and Phenomenal Knowledge: New essays on consciousness and physicalism*, Alter T, Walter S, (eds) pp 111-144. Oxford: Oxford University Press.

Paternoster A, (2014) Reconstructing (Phenomenal) Consciousness. In *Mind, Values, and Metaphysics: Philosophical Essays in Honor of Kevin Mulligan Vol.2*, Reboul A (ed) pp 249-260. Switzerland: Springer International.

Pereboom D, (2011) *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.

Pereboom D, (2014) Russellian Monism and Absolutely Intrinsic Properties. In *Current Controversies in Philosophy of Mind*, Kriegel U (ed) pp 40-69. London: Routledge.

Picciuto V, (2011) Addressing Higher-Order Misrepresentation with Quotational Thought. *Journal of Consciousness Studies* 18:109-136.

Premack D, Woodruff G, (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1: 515-526.

Priest G, (2014) *One: Being an Investigation into the Unity of Reality and of its Parts, including the Singular Object which is Nothingness*. New York: Oxford University Press.

Prinz J, (2016) Against illusionism. *Journal of Consciousness Studies* 23: 186-196.

Quine WVO, (1980) On what there is. In *From a Logical Point of View*, Quine WVO pp 1-19. Cambridge, MA: Harvard University Press.

Ramsey F, (1978) *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, Mellor DH (ed). London: Routledge.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L, (1996) Premotor cortex and the recognition of motor actions. *Brain Research. Cognitive Brain Research* 3: 131-141.

Robinson H, (1982) *Matter and Sense*. Cambridge: Cambridge University Press.

Rosenthal D, (1986) Two Concepts of Consciousness. *Philosophical Studies* 49: 329-359. Reprinted in *Consciousness and Mind*, Rosenthal D, (2005) pp 21-45. New York: Oxford University Press.

Rosenthal D, (1993a) Thinking that One Thinks. In *Consciousness: Psychological and Philosophical Essays*, Davies M, Humphreys G, (eds) pp 197-223. Oxford: Basil Blackwell. Reprinted in *Consciousness and Mind*, Rosenthal D, (2005) pp 46-70. New York: Oxford University Press.

Rosenthal D, (1993b) Higher-Order Thoughts and the Appendage Theory of Consciousness. *Philosophical Psychology* 6: 155-166.

Rosenthal D, (1997) A Theory of Consciousness. In *The Nature of Consciousness*, Block N, Flanagan O, Guzeldere G (eds) pp 729-754. Cambridge, MA: MIT Press.

Rosenthal D, (2004) Varieties of higher-order theory. In *Higher-Order Theories of Consciousness*, Gennaro, R (ed) pp 17-45. Philadelphia: John Benjamins.

Rosenthal D, (2005) *Consciousness and Mind*. New York: Oxford University Press.

Rosenthal D, (2011) Exaggerated reports: reply to Block. *Analysis* 71: 431-437.

Rosenthal D, (2015) Qualities Spaces and Sensory Modalities. In *Phenomenal Qualities*, Coates P, Coleman S (eds) pp 33-66. New York: Oxford University Press.

- Rosenthal D, (2018) Misrepresentation and Mental Appearance. *Transformação Revista de Filosofia* 41: 49-74.
- Rowlands M, (2001) Consciousness and Higher-Order Thoughts. *Mind & Language* 16: 290-310.
- Searle J, (1969) *Speech acts*. Cambridge: Cambridge University Press.
- Searle J, (2000) Consciousness. *Annual Review of Neuroscience* 23: 557-578
- Shoemaker S, (1968) Self-reference and self-awareness. *The Journal of Philosophy* 65: 555-567.
- Shoemaker S, (1979) Identity, Properties, and Causality. *Midwest Studies in Philosophy* 4: 321-42.
- Shoemaker S, (1980) Causality and Properties. In *Time and Cause: Essays Presented to Richard Taylor*, van Inwagen P (ed) pp 109-135. Dordrecht: Reidel.
- Siewert C, (1998) *The Significance of Consciousness*. New Jersey: Princeton University Press.
- Speaks J, (2015) *The Phenomenal and the Representational*. Oxford: Oxford University Press.
- Steward H, (1997) *The ontology of mind*. Oxford: Oxford University Press.
- Stoljar D, (2014) Four Kinds of Russellian Monism. In *Current Controversies in Philosophy of Mind*, Kriegel U (ed) pp 17-39. New York: Routledge.
- Strawson G, (2008) The Identity of the Categorical and the Dispositional. *Analysis* 68: 271-282.
- Swinburne R, (1980) Properties, Causation, and Projectibility: Reply to Shoemaker. In *Applications of Inductive Logic*, Cohen L, Hesse M, (eds) pp 313-320. Oxford: Oxford University Press.
- Taylor H, (2018) Powerful qualities and pure powers. *Philosophical Studies* 175:1423-1440.

Taylor H, (2020) The relation between subjects and their conscious experiences. *Philosophical Studies* 177: 3493–3507.

Thomasson A, (2000) After Brentano: A One-Level Theory of Consciousness. *European Journal of Philosophy* 8: 190-209.

Tye M, (1995), *Ten problems of consciousness*. Cambridge, MA: MIT Press.

Van Gulick R, (2000) Inward and Upward – Reflection, Introspection, and Self-Awareness. *Philosophical Topics* 28: 275-305.

Van Gulick R, (2004) Higher-Order Global States (HOGS). In *Higher-Order Theories of Consciousness*, Gennaro, R (ed) pp 17-45. Philadelphia: John Benjamins.

Van Gulick R, (2006) Mirror Mirror – Is That All? In *Self-Representational Approaches to Consciousness*, Kriegel U, Williford K, (eds) pp 11-40. Cambridge, MA: MIT Press.

Van Gulick R, (2022) Consciousness and Self-awareness – an Alternative Perspective. *Review of Philosophy and Psychology* 13: 329-340.

van Inwagen P, (1998) Modal Epistemology. *Philosophical Studies* 92: 67-84.

Voltolini A, (2016) Varieties of Cognitive Phenomenology. *Phenomenology and Mind* 10: 94-107.

Webb TW, Graziano M, (2015) The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology* 6: 1-11.

Weisberg J, (2008) Same Old, Same Old: The Same-Order Representation Theory of Consciousness and the Division of Phenomenal Labor. *Synthese* 160: 161-181.

Williams N, (2019), *The Powers Metaphysic*. New York: Oxford University Press.

Williford K, (2006) The Self-Representational Structure of Consciousness. In *Self-Representational Approaches to Consciousness*, Kriegel U, Williford K, (eds) pp 111-142. Cambridge, MA: MIT Press.

Wimmer H, Perner J, (1983) Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. *Cognition* 13: 103-128.

Wittgenstein L, (1958) *The Blue and Brown Books*. Oxford: Blackwell.


Worley S, (2003) Conceivability, Possibility and Physicalism. *Analysis* 63: 15-23.

Wu W, (2020) Is Vision for Action Unconscious? *Journal of Philosophy* 117: 413-433.

Zahavi D, (1999) *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwestern University Press.

Zahavi D, Kriegel U, (2016) For-Me-Ness: What It Is and What It Is Not. In *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*, Dahlstrom D, Elpidorou A, Hopp W, (eds) pp 36-53. New York: Routledge.

Zemach E, (1985) De Se and Descartes: A New Semantics for Indexicals. *Noûs*, 19: 181-204.

A handwritten signature in cursive script, appearing to read "David La Fatale", written above a horizontal dotted line.