Denis Delfitto / Gaetano Fiorin

Harmony Effects Across Language and Perception

Some Conundrums around the Unity of the Mind

PETER LANG

Denis Delfitto / Gaetano Fiorin

Harmony Effects Across Language and Perception

This book offers original insights around a fascinating idea: Perception and the rest of cognition, crucially including language, are closer to each other than the Cartesian tradition dared to dream. By combining recent results in cognitive neuroscience, the philosophy of perception, and the syntax of natural language, the book demonstrates that there is continuity between higher and lower cognition. Percepts from perceptual experience are propositional, conceptual, and they are not divorced from objective reality. Human cognition is merged with the natural world, able to reflect it in complex ways and interact with it in modalities that are since the very beginning computationally complex and rich in content.

The Authors

Denis Delfitto is Professor of Linguistics at the University of Verona. His publications span from the syntax of reference and quantification to issues of language change and language impairment. His core interest is the investigation of the human capacity for language from a broad cognitive and philosophical perspective.

Gaetano Fiorin is Assistant Professor of Linguistics and Logic at University College Utrecht and a researcher at the Utrecht Institute of Linguistics (UiL-OTS). His research spans from theoretical linguistics to experimental linguistics and the philosophy of language.

www.peterlang.com

Harmony Effects Across Language and Perception

STUDIES IN PHILOSOPHY OF LANGUAGE AND LINGUISTICS

Edited by Piotr Stalmaszczyk

VOLUME 17

Advisory Board:

Emma Borg (University of Reading)

Manuel García-Carpintero (University of Barcelona)

Hans-Johann Glock (University of Zurich)

Paul Livingston (University of New Mexico)

Joanna Odrowąż-Sypniewska (University of Warsaw)

Maciej Witek (University of Szczecin)

Marián Zouhar (Slovak Academy of Sciences, Bratislava)

PETER LANG

Denis Delfitto / Gaetano Fiorin

Harmony Effects Across Language and Perception

Some Conundrums around the Unity of the Mind

PETER LANG

This publication has been peer reviewed.

www.peterlang.com

Harmony Effects Across Language and Perception

To Paola and Linda,
Our fellow travelers
In the quest for harmony

Table of contents

Acknowledgments

The inquiries into the systems of language and perception reported in this book stem from the research project on linguistic meaning and the semantics of the first-person that led to Fiorin and Delfitto (2020). This is an intermediate phase and these are intermediate results: as we continue to be fascinated by the issues of meaning that arise at the confluence of cognition, perception and experience, our endeavor is set to continue.

For the stimulating intellectual exchanges we had during the process of writing this monograph, we have a deep intellectual debt to the following friends and colleagues: Alessandro Capone, Valentina Bianchi, Lisa Cheng, Norbert Corver, Martin Everaert, Alessandra Giorgi, Giuseppe Longobardi, Chiara Melloni, Jacques Moeschler, Andrea Moro, Giuseppe Varnier, and Maria Vender.

We are especially grateful to Eric Reuland, for his detailed comments on large parts of this manuscript and his generous encouragement, and to Anne Reboul, whose challenging philosophical suggestions were the original stimulus to some of the research endeavors reported here.

Last but not least, we are grateful to Piotr Stalmaszczyk and Lukasz Galecki (Peter Lang) for their precious editorial support, and to two anonymous reviewers for their helpful suggestions.

It goes without saying, usual disclaimers apply.

Introduction

The goal of this monograph is to offer some arguments in favor of the conceptual and propositional nature of perception. It argues against the Cartesian view of perception as a process of passive sensory registration. There is no 'veil of ideas' between the senses and the world. In fact, showing that the senses are already pervaded by the operations of reason is the best way to show that the world is not as dichotomic as an important part of our philosophical tradition would have it. There are not subjective sense-data opposed to an objective reality. There are simply biological organisms merged with the natural world, able to reflect it in complex ways and interact with it in modalities that are since the very beginning (that is, starting with the apparently most elementary operations of sensory registration) computationally complex and rich in content.

Cognitive neuroscience made us acquainted with the idea that perception and action are not divorced from the higher operations of cognition. If hearing, seeing, and acting hinge on complex processes of information extraction from analogical databases, it is then uncontroversial that they are cognitive processes. In fact, the embodiment movement (Varela et al. 1991) adds the suggestion that the traditional view of cognition as essentially *abstract* might be based on the wrong assumption that conceptual knowledge cannot be modality-specific. Embodied cognition is based on the insight that the necessary and possibly sufficient condition for cognition is the way in which sensorimotor capacities enable organisms to successfully interact with the environment. It follows that there is no independent need for abstract conceptual representations: these representations can be replaced by distributed networks of modality-specific sensory and motor features. Under this view, concepts are not abstracted away from sensory-motor experiences; concepts are rather identified with the process of reactivation of modality-specific experiences stored in the sensory-motor cortices (Pulvermüller 1999, Gallese and Lakoff 2005). Most typically, embodied cognition meets the mirror neurons hype, as when concepts of actions are identified with the motor patterns activated by perceiving an action. However, it has also been claimed – rightly so, we believe – that grounding cognition in sensory and motor systems is nothing else than

grounding cognition in cognition, since sensation is no less computational than abstract cognition:

> The output of the neuron is not a copy of its inputs. Instead its output reflects a weighted integration of its inputs. It is performing a transformation of the neural signals it receives. Neurons compute. This is information processing and it is happening in every single neuron and in every neural process whether sensory, motor, or *cognitive* (Hickok 2014).

In this monograph we support the view that perception is cognition. In doing so, however, we do not feel compelled to adhere to the embodied cognition view that there are no abstract representations of concepts in the brain. There is substantial evidence to the effect that (i) abstract *amodal* categories exist; (ii) they are actually encoded in the anterior temporal lobe; and (iii) they are justified in terms of evolutionarily-induced domain specificity (Caramazza and Mahon 2003, 2006).

What this monograph specifically emphasizes is that the lesson that perception is cognition has hardly been assimilated. The idea still resists that higher cognition is something entirely different from perception: contrary to perception, it is representational, conceptual, combinatorial, compositional, and propositional. Moreover, it is common to think that higher cognition comes in as a snowball effect: being endowed with an abstract concept means that the latter can be made itself an object of reflection; attributing a property to an object in a proposition entails that that proposition can be made itself a subject of predication. This wrong attitude is, in a nutshell, what Burge (2010) calls 'individual representationalism,' the tendency to believe that the capacity to entertain a given representation must be grounded in the awareness that we are entertaining that representation and in the capacity of using that representation as an ingredient of higher-order representations.

Individual representationalism, however, is wrong: the fact that a system works according to a certain set of laws does not necessarily entail that these laws must be represented within the system. Our mind works according to distinct sets of computational principles represented in different cognitive modules, and we humans, on a par with all other species, are only minimally aware of these computational principles.

It is also generally believed that higher cognition is based on properties that are not shared with perception. For instance, reference to objects and

events is established in language, because it involves awareness of what objects and events are, an awareness that only arises in language. In vision – it is argued – we do not necessarily see objects *as something*, as for instance in Dretske's notion of 'simple seeing' (Dretske 1981) and, when we do so, the suspect arises that this achievement might be an effect of cognitive penetration – the capacity of a perceptual system to compute functions that are 'sensitive, in a semantically coherent way, to the organism's goals and beliefs' (Pylyshyn 1999, p. 343). However, this is a typical fallacy of the intellectual attitude that denies continuity between perception and the systems of language and thought. This monograph takes issue with this stand, arguing that some of the properties that are typically considered as a prerogative of the systems of language and thought are already found in perception, and that they are in fact no less defining of perception than they are of language.

We aim at providing new original insights around a fascinating idea: perception and cognition are in fact much closer to each other than the Cartesian tradition dared to dream. The mind is more unitary than we usually imagine, in the sense that the alleged systems of 'lower cognition' *anticipate* some of the properties that are usually regarded as proper to higher cognition, sometimes in the sense that higher cognition actually *mirrors* properties that are typically encountered in perception.

These insights are also relevant to assessing and potentially resolving the opposition between externalism and internalism in semantic theorizing. This opposition has profoundly affected much of the discussion of linguistic meaning of the past century, particularly within the analytic tradition. On the one hand, semantic externalism, typically supported by twin experiments such as Putnam's famous twin earth thought experiment (Putnam 1975), maintains that linguistic meaning is a worldly object, whose identity conditions are to be found in the external world of material things and not in the inner psychological lives of language users. On the other hand, semantic internalism, a position most typically associated with the work of Chomsky (see, among others, Chomsky and Smith 2000), sets its empirical focus exclusively on speakers' linguistic competence – the inner states that characterize their capacity for meaningful language. Both these approaches, at least when taken in their strongest versions, have proven to be too restrictive and, ultimately, unsatisfactory. Many today support

more modest claims, such as the weaker externalist claim that at least *some* of the meaning of *some* linguistic expressions is identified externally to the minds of its users or the weaker internalist claim that, even though language undeniably allows speakers to refer to the external world, the goal of linguistic theorizing should nonetheless be restricted to the internal states that allow speakers to do so. However formulated, these approaches beg the question of how the worldly, public dimension of meaning can be reconciled with its private, cognitive significance. This challenge is typically addressed by heavily relying on two-dimensional logical frameworks such as Kaplan's two-dimensional semantics of demonstratives and indexicals (Kaplan 1989). Whereas it remains unclear whether similar approaches are in fact successful, the more radical solution we would like to put forward in this book consists in questioning the very divide between internalism and externalism in semantic theorizing. As we have mentioned, and will discuss at length in the coming pages, a large body of empirical work performed in recent decades in the domain of perception has proven the Cartesian divide between world and mind to be but a profoundly mistaken theoretical stand. It is only by learning this lesson that we can objectively reconstruct the divide between cognitive agents and their ecology and, in this way, aspire to a unified notion of linguistic meaning.

Quite importantly, this book does not aim at providing an empirical description of how language reflects perception – an endeavor that falls within the scope of much work in contemporary cognitive linguistics. In particular, a long tradition of research in cognitive linguistics has argued, in a very elaborate way, that language is strictly intertwined with perception, and that many aspects of the way in which meaning is expressed in the syntax of language should be analyzed as the natural reflection of how perceptual processes unfold (Lackoff 1987, Gallese and Lakoff 2005, Langacker 1987). Similarly, an influential tradition of research in conceptual semantics (Landau and Jackendoff 1993, Jackendoff 2002) has investigated how spatial functions are encoded by spatial prepositions in language, and, more generally, to which extent linguistic categories reflect the spatial and geometric properties of the entities (objects and places) that fall within their domain of reference, suggesting that the linguistic divide between objects and places (the 'what' and the 'where') follows from a more basic *perceptual* disparity. This long tradition of linguistic analysis has been

recently enriched by an already impressive series of brain-imaging studies and behavioral experiments suggesting that meaning arises, in language as well as in perception, not through the manipulation of abstract amodal symbols within a set of equally abstract computations, but rather through the simulation of actions and perceptions (Barsalou 1999, Vulchanova et al. 2019). By extending some of the basic insights of the stream of research in embodied cognition, the Visual World Paradigm introduced dedicated experimental methods aiming at an in-depth investigation of the language-perception relationship, focusing particularly on the interaction with vision. According to this view and paradigm of research, perceptual information crucially shapes the interpretation of linguistic stimuli, to the effect that even the meaning of what is generally considered logical parts of discourse, such as negation, is radically refashioned in terms of 'experiential' processes that visually simulate the negated state-of-affairs as part of the procedure for 'interpreting' negation (Kaup 2007, Vender & Delfitto 2010, Scappini et al. 2016).

As for us, we definitely think that the substantial interplay between the systems of language and perception cannot be reduced to the thesis that linguistic meaning arises and develops in terms of processes based on the simulation of perceptual states. As a matter of fact, in Fiorin and Delfitto (2020) we offered a completely different interpretation of the thesis of 'meaning as perception.' We endorsed the view according to which the things we talk about in natural language behave *as if* they were objects of perception. On the one hand, we emphasized the fact that 'the parallels between the logical structure of perception – as uncovered through the scientific study of sensory systems – and the logical structure of linguistic meaning – as envisaged by model-theoretic semanticists – are remarkable' (Fiorin and Delfitto 2020: 309). On the other hand, we argued that the metaphysical properties of natural language meaning emerge more clearly when one explicitly draws the parallelism with the metaphysics of perception, which should be conceived, pretty much along Tyler Burge's lines, as establishing the roots of objectivity. In our view, it is this shared metaphysics that offers the most valuable framework for reconciling the roles played by world and mind in determining the nature and the properties of meaning. In the present book, we pursue and extend this line of inquiry, offering some further investigations into the 'harmony effects' that

surprisingly arise when we carefully assess the properties of the systems of language and perception.

To put it shortly, this book intends to shed light on what language and perception have in common, and on how precious these commonalities are when we try to elucidate the issue of meaning in human cognition. However, it firmly objects to the thesis that linguistic meaning can be reduced to perceptual processes. In fact, we do not intend to support a usage-based view of language that reduces the abstract conditions that apply to language structure and its interpretation to the pragmatic conditions that apply to sensorial perception. Such a conclusion would not do justice to the rich view of perception that we will defend in the next chapters and, in any case, would require a discussion of its own, which, if disclosed in all the necessary details, would take us too far away from the goals we have set for ourselves in this book. However, the reader interested in the broader philosophical pragmatism about language that emerges from the views proposed in this book is referred to Fiorin and Delfitto 2020 (cf. in particular Parts 2 and 3).

In its search for harmony across the cognitive domain, this monograph does not underestimate what sets apart language and perception. On the contrary, it is based on the understanding that the bootstrapping from modality-specific perceptual combinatorics (as in the complex spatio-temporal dynamics associated with the visual representation of a transitive action involving an agent and an undergoer) to logical/linguistic combinatorics involving amodal features like the thematic roles of agent and undergoer (and more generally argument structure) represents a major cognitive leap forward. A leap forward that is made even more robust by the activation, in language, of rigid templatic structures that regiment the higher-order logical space associated with verbs and nouns (Cinque 2002; Borer 2005a/b, 2013). This is the so-called functional structure of language, a basic ingredient of syntax. How are these abstract cognitive properties exactly represented in the brain? Presently, we do not know, and this monograph does not add to our knowledge of these fundamental issues.

As emphasized above, we have also nothing to add to the tradition of studies in conceptual semantics and cognitive linguistics that is set to investigate how perceptual features are reflected in linguistic structure, apart from the inquiries into the semantics of 'verbs of seeing' presented in Chapter 5.

Moreover, this monograph is not conceived as a typological study of how (lexical) semantics reflects perceptual distinctions across the languages of the world and how different languages set up the trade-off between sensory data and epistemic states of belief and justified belief (knowledge). In fact, the inquiries reported in this monograph are limited to English and a few other languages. Also, there are presently a number of interesting suggestions revolving around the issue of how language expresses primary sense qualities like hearing, touch, smell and taste (see, for instance, Krifka 2010 and the references cited therein). We are convinced that these inquiries may be directly relevant for the purposes of this monograph. However, the task we set ourselves here is not that of inquiring into the ways in which the architecture of the mind or specific cultural experiences impinge on the linguistic coding of sense impressions. Once again, our task here is revealing the presence of a common set of abstract conditions on representation, which undermines or at least considerably weakens the traditional divide between sensory impressions and perceptual representations, on the one hand, and the systems of thought and reason, on the other.

Ultimately, this monograph has the ambition to put forward some new arguments for the unity of the mind. If some of these arguments stand, it is this unity that emerges as one of the most bewildering properties of the mind.

The book is organized as follows. Chapter 1 develops the main insight of the book by presenting a number of arguments against the view that the systems of perception involve a serious epistemological divide with respect to the systems of language and thought. In order to argue against this divide, and in favor of a non-negligible intersection between the abstract properties of perception and language, we elaborate on the tradition of studies in the philosophy of mind according to which perceptual acts have a content and this content is necessarily representational and propositional. In this way, the representational format, together with the veridicality conditions and requirements that are traditionally associated with it, extends across all the systems of cognition, from vision and audition to memory, language and thought. Propositional knowledge and veridicality conditions are not an absolute prerogative of language; in fact, perception is not only – at least to a certain extent – combinatorial and compositional but in fact also, and crucially, propositional.

Chapter 2 is a philosophical interlude whose goal is to provide the reader with a background discussion of the literature on the philosophy of perception, with a specific focus on the controversial notion of perceptual content. Though there is obviously no pretension of exhaustiveness in this presentation of the state-of-the-art, this part offers a preliminary review of the relationship among percepts, concepts, representations and propositions, showing how these notions play a central role in the assessment of how perception relates to cognition.

Chapter 3 expands the insight that the conceptual/propositional representations that are considered as a prerogative of language are already an essential property of visual representations. More specifically, the capacity for reference to objects and events that is typically attributed to language is already shared by the visual processes that lay the foundation of knowledge and action. We further argue that the modes of reference detected in language (*descriptive* and *direct* reference) are the very same modes of reference that are detected in vision.

Chapter 4 shows that the semantics of verbs of vision mirrors some non-trivial ingredients of a serious epistemology of perception. The divide between epistemic seeing (whereby an act of seeing is interpreted as associated with a state of knowledge) and perceptual seeing (whereby an act of seeing is interpreted as supporting a state of belief) is encoded in language. More precisely, we argue that what is encoded is the divide between *epistemic* seeing and what we will call, introducing an important distinction, *experiential* seeing, suggesting – a quite astonishing finding indeed – that the systems of language already incorporate an 'implicit' model of how vision supports belief and feeds knowledge.

Finally, Chapter 5 develops the further insight that the syntactic subcategorization properties of verbs of propositional attitude, and more specifically the presence of a 'silent' event noun as part of their argument structure, surprisingly mirror a specific epistemological stand towards propositions. More precisely, there seem to be properties of the syntactic representations that suggest that propositions are cognitive events of property-attribution, a position explicitly endorsed by Scott Soames (King at al. 2014, Soames 2015). This finding challenges the divide between

descriptive and foundational metaphysics. Contrary to what is commonly assumed, the objects that provide the basic ingredients of language-driven interpretive processes (objects, events, facts, etc.) may have – at least in some cases – an epistemological significance that is not strictly confined within the language module.

Chapter 1   Two views on language and perception

1.1 Language and individual representationalism

Within the mainstream philosophical traditions that T. Burge polemically dubbed 'individual representationalism' (Burge 2010), it is widely held that the possibility for an individual to generate representational states applying attributes to objects in such a way that this attribution represents naturalistically relevant aspects of reality depends, in each individual, on a higher-order capacity to represent the explicit conditions under which this objective representation becomes possible. These conditions are commonly thought of as embodying the capacity to represent to oneself the difference between appearance and reality, as well as the capacity to conceive of causal principles governing the interaction between external objects and a mind engaged in acts of perception and cognition. For instance, it is generally believed that one cannot entertain the proposition 'that object o is red,' in perception or cognition, without being able to entertain the concept of 'proposition' itself, and without being able to use this concept in higher-level representations in which, say, attributes are applied to propositions.

It is perhaps not so far-fetched to imagine a honey bee being able to entertain a perceptual representation in which the attribute 'yellow' is applied to a particular flower. In reality, it is likely that the bee's visual system is not empowered with this sort of perceptual content. However, we can easily *imagine*, after all, that the bee is wired in such a way that it is this representation of the flower as 'yellow' that causes the bee to go to the flower (whereas, say, it is the representation of a flower as 'red' that causes the bee to get away from the flower). Suppose this is the case, for the argument's sake. Would we accept that the capacity of the bee to create a *visual* representation in which object o (the flower) has property Y (being yellow) is formally equivalent to the capacity humans have to express the same perceptual representation propositionally in language, by uttering, say, the sentence 'that flower is yellow'?

The customary answer is a vehement 'no.' The reason is that it is thought that when we entertain proposition p by expressing sentence S (say, 'that

flower is yellow') we possess the further capacities of comparing the concept of *flower* and the concept of *yellow* and of using these concepts in independent representations, like 'flowers are beautiful,' and in higher-level representations, of the sort of 'yellow is a color' or 'whatever is yellow is useful.' Hence, the capacity to entertain an elementary representation such as *'that flower is yellow'* must be rooted in a far more complex capacity to freely use the relevant concepts in ways that make these concepts themselves the objects of our representations.

This view has at least two important consequences, which we will discuss in some detail. We shall see that, ultimately, neither of them is sustained. Individual representationalism has it wrong.

1.1.1 The first consequence of individual representationalism

Individual representationalism entails that our capacity to refer to objects, events and facts in language must be rooted in our capacity to epistemically distinguish between objects, events and facts, as when one says 'I saw my baby smile,' intending that she witnessed a visual event in which her baby smiled, or when one says 'I regret that you left,' intending that your leaving is a fact and that she regrets that fact.

We refer to facts and events, in language, as formal entities with distinctive properties. It is part of our ability to use language in a way that presupposes the existence of objects of a certain sort. Philosophers and linguists have traditionally used linguistic data to show that we are in fact implicitly committed not only to the existence of commonsense things such as material objects, artifacts, events and facts, but also to the existence of weird things such as holes and tropes. This is commonly referred to as 'descriptive metaphysics' (Moltmann 2013). A part of what we mean when we utter a sentence in a language we speak is undoubtedly due to the implicit ontology that is associated with that language. For instance, if one told you 'Yesterday, I saw my baby smile but in fact it was not my baby,' you would probably conclude that she is reporting an optical illusion of some sort. But if one told you 'I regret that you left, though you didn't leave' or 'I saw that my baby smiled, but in fact it was not my baby,' you would feel entitled to conclude that she must have gone mad, since she expressed contradictory contents. These different judgments are arguably rooted in your knowledge that certain clause-types refer to a particular sort

of events (visual experiences), whereas other clause-types refer to facts, as well as in the concomitant knowledge that facts and events have distinct metaphysical properties.

As we will see in Chapter 4 in more detail, the different conclusions on one's communicative intentions and one's mental state depend on the properties of the abstract objects whose existence language presupposes. By uttering the sentence 'I saw my baby smile,' the speaker is reporting her own visual experience. If it turns out that the baby she saw was in fact not her own, we are bound to conclude that her visual experience does not correspond to a *correct* perceptual act: in her experience, she actually saw her baby, though in reality it was someone else's baby. Conversely, by uttering the sentence 'I saw that my baby smiled,' the speaker is reporting that by being exposed to a certain visual scene, she acquired knowledge that her baby smiled. This knowledge is justified by the visual experience she had; the existence of the visual experience in which her baby was smiling is thus automatically presupposed and cannot therefore be denied by adding the information that the smiling baby turned out not to be hers. If someone talked like this, we would think that she is confused, since what she said is simply contradictory. We would not think that she is reporting an illusion or hallucination of some sort. These are just hard facts about knowledge of language. Notice further that, of course, it is not contradictory to say 'I believed that my baby smiled, but it turned out it was not my baby.' This is hardly surprising, given that there is a solid body of philosophical knowledge informing us that 'believing that p' is epistemically weaker than 'knowing that p,' since the former, but not the latter, is compatible with the falsity of p. More interestingly, the observation that 'I saw my baby smile, but it wasn't my baby' is not necessarily a contradiction suggests that this kind of perceptual report (where the complement of 'see' is a naked infinitival) connects to belief, and not to knowledge, semantically and epistemically.

Here are thus the facts. One can use the sentence 'I saw my baby smile' to report a hallucination (there was no child smiling) or some sort of optical illusion (the baby was not *her* baby), but one cannot use the sentence 'I saw that my baby smiled' to consciously report some hallucination or optical illusion she experienced (unless one decides, of course, to play with language to express mad thoughts). The reason is that when one employs the latter

sentence one reports on the knowledge she has acquired that her baby was smiling, and hallucinations or optical illusions are not, of course, reliable sources of knowledge.

This observation takes us to the central question: Are speakers *consciously aware* of the distinction between these two types of complementation structures ('see that,' followed by a finite clause and 'see' followed by a bare infinitival)? And more generally, are they consciously aware of the epistemological difference between reports of *visual perceptions* (reporting mental states that feed belief) and reports of *visually justified states of knowledge* (reporting perceptually-acquired states of knowledge)? Hardly so. Despite a lot of work by distinguished philosophers and cognitive scientists, there is still considerable disagreement among scholars themselves about how the concepts of experience, perception, sensation, belief and knowledge are epistemologically related to each other. So, there is hardly any doubt that when one entertains a representation involving an abstract fact and not a perceived event by uttering 'I saw that my baby smiled,' one cannot be easily held to be *consciously* doing that. This entails that either these representations – involving facts, events, and the formal difference between them, whatever they are – do not exist as such (but then, why do they influence one's behavior, by making one think either that the speaker is talking nonsense or that she had a hallucination?) or that one is in fact using, when she speaks, reference to facts, events and other weird objects of this sort, but does that *unconsciously*, that is, without being aware of the ontological distinctions that are not only *implicitly* made but also part and parcel of the mental processes by means of which one draws inferences and influences behavior.

If one accepts the latter conclusion, then one should also accept that having complex conscious thoughts about perceptual representations or abstract objects such as propositions is *not* a necessary condition for someone to entertain perceptual representations or propositions, after all. Similarly, one should accept that making an *implicit* use of the ontological differences between facts and events (or between belief and knowledge in the use of propositional attitudes) does not compel her to possess any *explicit* knowledge of what these differences exactly entail in epistemological terms. Being an outstanding philosopher is in no way a necessary condition for being a proficient language user. The fact that one has

complex representations and makes complex computations does not entail that one must be aware of the existence or nature of these representations and computations. The need for higher-order representations is a myth. Individual representationalism has it wrong, T. Burge is completely right in forcefully making this point.

We conclude that individuals can build, when they use language, representations involving abstract objects of distinct sorts without having *explicit* concepts of these objects, that is, without being able to turn these objects and their properties into the components of higher-level representations they may entertain. Philosophers, linguists and cognitive scientists may do that, but it would be ludicrous to propose that *only them* can utter sentences like 'I saw that my baby smiled' and 'I saw my baby smile' while assigning to them the distinct meanings (and the causal powers on behavior) that we have discussed above.

1.1.2 The second consequence of individual representationalism

The idea that propositional thought comes in as a whole (we cannot entertain the proposition that 'that flower is yellow' unless we are able to entertain the proposition that 'what is yellow is useful' or things like that) entails that a languageless infant, whenever she produces a perceptual representation of *that* flower being yellow, cannot be really assumed to entertain the 'proposition' that that flower is yellow. The reason is that we cannot grant the infant with the possession of the concepts 'yellow' and 'flower,' since, by definition, possessing them would turn the infant into a philosopher capable of entertaining the propositional thoughts that, say, flowers are not animals and that not every yellow thing is a flower. We would then be forced to conclude that the infant lives in a world of pre-conceptual representations in which she is endowed with a sort of primitive surrogate of the visual representation of the flower being yellow, in which a yellow flower out there causally produces, in the infant, an informational state that may impinge on behavior, without the infant being able to use that informational state for further computations.[1]

---

1   Such a pre-conceptual state would be comparable – as it is sometimes

Be it as it may, we should also be aware of a second, and perhaps more serious, problem with denying propositional representations to infants. It is that, if we do so, we should also deny propositional representations to adults that are not capable of handling the concepts of event and fact when they utter sentences of the sort of 'I saw my baby smile,' in which they *implicitly* refer to events or facts. If we accept that we can manage, in language, facts and events without conscious knowledge of what they are, we should also accept that we can build propositional representations even at stages (such as in perception) in which we do not dispose of conscious knowledge of what propositions are.

1.1.3 Perception, cognition and grammar

From these observations, we conclude that 'individual representationalism' is a philosophical hallucination produced by a widespread fallacy, consisting in the idea that thoughts are like snowflakes, and that there are no things like simple snowflakes and complex snowflakes. More particularly, thought comes in with its full apparatus of objects, properties, properties of properties (what we call, in language, 'quantifiers'), events, propositions, facts, tropes and so on. Those who are not capable of producing mental representations where properties are the objects of which other properties are predicated (as when we say 'everyone is happy') or mental representations in which propositions are the objects of which properties are predicated, as when we say that 'everything you thought is false,' cannot be granted the possession of propositional states representing what is out there in terms of sentences as simple as 'that is yellow.'

---

suggested – to the pre-conceptual conditions in which humans find themselves when they are subject to *simple seeing*, in F. Dretske's sense (see Dretske 2000), as when, for instance, one sees *something*, in the sense that one visually perceives its presence, but cannot say what kind of object it is or even, conceivably, which color or shape it has. It is a seeing deprived of one of the typical higher-level properties of seeing, *seeing as*. We see something without being *conceptually* aware of what we are seeing. We will argue in Chapter 2 that non-conceptual seeing uncontroversially exists. In fact, non-descriptive seeing is a central aspect of *adult* human vision (it is a feature of Pylyshyn's FINSTs, as well a feature of motion-directed vision, see Milner and Goodale 2006). It is certainly not the alleged prerogative of languageless infants' vision.

As we have seen, this a fallacy, and a fallacy with many unpleasant consequences, though it is a difficult one to eradicate, not least because it gives us the comforting assurance that only we *humans* are capable of thinking about the world, in the sense of creating objective representations of what there is as it is. Nevertheless, it is a fallacy, and a serious one. We concur with Burge that:

> Deeper understanding of the failure of Individual Representationalism derives from reflecting on science, particularly perceptual psychology, developmental psychology, and ethology. Some of this science matured only since the 1970s […] Throughout the century, empirical work on perception had astonishingly little impact on philosophical reflection on perception. A better account of origins of objective representation must center on better understanding of perception. Elaboration of anti-individualism about perception and appreciation of the science of perception not only help undermine Individual Representationalism. They point toward a different understanding of origins of objectivity. (Burge 2010: 107–108)

This 'different understanding,' we believe, should be based on the acknowledgement that concepts, as well as propositions, can be successfully entertained even though we may not be aware of the fact that they are concepts and propositions, that is, abstract objects endowed with properties that could be ascribed to them at a higher level of representation (Fodor and Pylyshyn 2015). This boils down to acknowledging that complex representations of a propositional nature are not the exclusive prerogative of higher-level cognitive systems such as language. In the words of Scott Soames,

> How we see things — the predication we make — is usually automatic, unconscious, and so better described as a kind of cognitive operation than as a species of intentional action. But sometimes our experience makes multiple properties or relations available for predicating of the same things, either unconsciously or with a degree of conscious control—in which case our predications occasionally qualify as intentional. Either way, the important point is that putting together representational structures in perception and cognition is always a cognitive operation of some kind. The simplest cases are those in which we predicate properties or relations of things that are given to us in perception or cognition, and thereby entertain a simple proposition, like the proposition that o is red, or that o_1 is bigger than o_2. (King et al. 2014: 95)

Propositional representations are, within certain limits, given in perception no less than in cognition and a large part of the thoughts we entertain

when we express these thoughts in language involves the manipulation of concepts of which we are not more conscious than a honey bee is of the concept of flower whenever she undergoes the phenomenal experience that a certain flower is yellow.

As we will see in the next chapters, a correct understanding of propositional representations involves getting rid of the fallacy that language allows us to think *consciously*, at least if we interpret 'thinking consciously' as consciously manipulating all the concepts involved in the linguistic expression of a thought. In fact, we will see that there are suggestions even to the effect that meaning is experienced as a monadic property of linguistic expressions: we are simply not conscious of the (contextual) ingredients we are assembling together when we undergo the phenomenological experience of associating a meaning to a sentence to which we are exposed (Azzouni 2013). Grasping the meaning of a sentence might be less similar to consciously playing around with words, concepts, gestures, intentions and so on, than the Gricean and neo-Gricean tradition would want us to think. It would be in a sense more like perceiving a linguistic expression from a certain perspective, which is in turn made available by a complex array of totally unconscious perceptual and cognitive dispositions. The representations we build of what there is as it is are not the product of our conscious individual effort. Rather, they are the product of complex computations – performed in distinct subsystems of perception and cognition – which originate from a series of long and tangled phylogenetic processes. The point to emphasize is that, in principle, the representational expressive power these computations have in perception is not inherently different from the expressive power they have in language and thought. The idea that what an infant does when she builds up the visual representation of a certain flower as yellow should be described in terms of passive 'sensations' and mechanical processes of stimulus-response is probably not less wrong than the idea that *an adult's* visual representation of the same flower as yellow falls short of the properties of intentional reference and explicit concept manipulation that characterize the proposition 'that flower is yellow' whenever it is expressed linguistically.

Eventually, we should recognize that there are simple snowflakes and complex snowflakes. Though not everything we get in cognition can be found in perception, since these systems allow a different degree of

combinatoriality and compositionality, the representational nature of perceptual systems is essentially the same as the representational nature of higher-level cognitive systems such as language.

In the Chomskyan tradition, language is often regarded as a cognitive system that is not in continuity with the systems of perception, in the sense that it has new and unprecedented properties. There is, additionally, a potentially misleading correspondence between the idea, proper to individual representationalism, that there are no such things as proto-concepts or a proto-thought, and the idea that there is no such thing as a proto-language. This idea is effectively presented in the following fragment from A. Moro, in which he also quotes Chomsky.

> First, any sentence utilizes, in a sense, the entire structure of grammar, just as any arithmetical expression utilizes the entire structure of arithmetic. Second, there can be no proto-syntax since the core notion of syntax by definition involves infinity, and there is no such thing as proto-infinity. Third, there is no room for evolution, but there certainly is for change; snowflakes, after all, do not have a history — they may all be different, but no single one of them is simpler than any other or the ancestor of any other. In this sense, it is language as a whole that is manifested in the structure of each and every sentence, and we can also regard a human language as an upper-dimensional snowflake, or perhaps a snowfall. This has also an interesting meaning when it comes to evolutionary perspectives, of course as far as syntax is concerned. To quote Noam Chomsky, 'Language is more like a snowflake than a giraffe's neck. Its specific properties are determined by laws of nature; they have not developed through the accumulation of historical accidents' […]. Here — as it seems to me — we face one of the most striking and destabilizing paradoxes of nature and surely the one that pertains to us most: a finite object shaped by evolution (the brain) expresses a code that generates infinite discrete structures and that cannot evolve by definition (syntax). Once more, human language reveals itself as the constant scandal of nature. (Moro 2016: 28–29)

Now, the point is that though some aspects of the issues raised in this quote still remain largely elusive, Moro is probably right in contending that what Chomsky defined as the faculty of language in the *narrow* sense came in as a snowflake (that is, as a whole) rather than as a progressively evolving system. It is hard to conceive of properties like discrete infinity, recursion, and locality as evolving in distinct stages: a computational system either has or has not them, and that's all. However, as we shall see in a moment, this idea is in danger of being confused with other ideas that are just plainly wrong: the idea that language is what enables conceptual/propositional

thought, the idea that language is completely detached from perception and, what is even more seriously flawed, the idea that there cannot be an objective representation of the 'world' without grammar.

There is no doubt – we believe – that this is not Chomsky's position. There are properties of syntax, like those mentioned above, that distinguish syntax, as a formal object, from the formal properties that can be reasonably ascribed to perceptual representations. And cognition – not only language – has properties that significantly extend beyond the properties of perception (*strong decoupling*,[2] in the sense of Reboul 2017, and *quantification*,[3] as in Soames 2015, are just two examples, among others). However, this observation should not be confused with the claim that the faculty of language in the *broad* sense is completely detached from perception or with the even more dangerous – and fallacious – claim that syntax is the only cognitive foundation of the human capacity of representing the world objectively. In what follows, we will examine the nature of these fallacious claims, which are based on a serious neglect of the results of the scientific inquiry in perceptual psychology and ethology.

Maybe syntax – to use Moro's words – is a scandal of nature. In fact, in the next chapters we will argue that there are additional reasons – actually strictly connected to the results of the discussion in the present chapter – to corroborate this view, however disturbing it may be to the Darwinian canon. But language – here intended as the faculty of language in the broad sense (Hauser, Chomsky and Fitch 2002) – certainly is not. There is no doubt that language astonishingly extends the representational power of perception, as does cognition (thousands and thousands of the abstract concepts that we entertain and manipulate in thought are *not* perceptual concepts). But it is certainly not syntax – and it is certainly not high-level

---

2   This is the capacity of entertaining representations in the absence of their content (the objects these representations are representations of) and in a way which is not directed towards action.

3   Here, we conceive of quantifiers in accordance with the Fregean tradition, that is, as involving representations in which a second-level property is predicated of a first-level property. So, when we say, for instance, that 'everyone runs,' the thought we express is that the first-level property of running satisfies the second-level property according to which running is instantiated by every object in the domain of discourse.

cognitive systems – that laid out the foundations for the objective representation of the world. Rather, these foundations lie in the workings of perceptual systems (Fiorin and Delfitto 2020).

1.1.4 Representing the world in language and perception

In order to be aware of the intimate connection between language and the systems of perception, it may be useful to pursue a program that – somehow paradoxically – intellectualizes perception and de-intellectualizes language. As is well-known, the philosophical debate about perceptual content, whether there is content and what its nature is, is quite alive today. In spite of the many, and often at least partially conflicting, positions that are taken on the issue (see next chapter for a full discussion), we think it is fair to say that perception is quite more conceptual and propositional in nature than is assumed in some linguistic circles. And, as we shall see in Chapter 3, the representational properties of language – crucially including its referential properties – are quite less detached from the referential and representational properties of perception than it is commonly maintained.

The tendency to downplay perception and overestimate the role of language in the analysis of the epistemological conditions for objectification is perhaps most typically represented in some recent work by W. Hinzen (Hinzen 2014 and Hinzen & Sheehan 2013). Hinzen's view of the relation among language, cognition and perception involves the claims that:

i. Intentional reference is based on the shift from perception to language;
ii. Reference is an exclusively linguistic business;
iii. There are concepts only when there are *lexemes*, since it is lexemes that establish the possibility of intensional and intentional reference;
iv. Concepts exist only when one is conscious of the thoughts that contain them (in the sense that she is capable, for instance, of verbally reporting them);
v. Only language creates the formal ontology that lays out the cognitive foundations for concepts: objects, actions, properties, propositions, events.

Notice, first of all, that according to Hinzen the emphasis is on the fact that the existence of pre-linguistic concepts being combined in a language of thought, somehow duplicating the combinatorial properties of syntax

when it applies to words/morphemes, is a philosophical chimera to say the least. Concepts *are* words, and the language of thought is syntax proper. Here are some telling quotes from Hinzen:

> In particular, as Carey stresses, conceptualization in non-linguistic beings is still continuous with perception; it also remains stimulus-controlled, non-combinatorial, and non-propositional, and concepts are not employed for purposes of intentional reference, with a capability to refer to anything at all no matter how remote in space and time (Fitch 2010: 187–194). (Hinzen 2014: 231)

And even more explicitly:

> We have referred to the term 'concept,' but generally concluded that what we were really talking about were either percepts (or at least entities continuous with percepts), mental representations, or lexemes. If so, the notion of a 'concept' is of little use, no case for prelinguistic concepts distinct from percepts remains, and we arrive at what is generally considered a 'harsh' stance on the animals. Our stance effectively assumes the existence of concepts only where there are lexemes, and indeed only where such lexemes enter into acts of intentional and intensional reference, which in turn require grammar. […] we will use the term 'concept' in what follows with this meaning: there are concepts in this sense only where there are lexemes, there is grammar, and there is intentional and intensional reference. This is our answer to the question of what 'concepts' really are, who has them, and what their semantics is. (Hinzen & Sheehan 2013: 54)

Downgrading percepts to non-concepts is not only equivalent to downplaying the representational and objectifying power of perception, it is also equivalent to completely subsuming the 'snowball metaphor': one cannot entertain the concept of 'flower' or of 'yellow' without being able to entertain the concepts of 'concept,' of 'proposition,' of 'property' and so on. As we have seen, this is the error of 'individual representationalism' in Burge's sense, and it amounts to asserting that thought, as syntax in the narrow sense, comes in essentially as a snowflake: either you take it as a whole or you don't take it at all.

Here is how Hinzen clearly expresses this perspective on the impossibility of decomposing propositional thought into a hierarchy of concepts and a hierarchy of distinct propositional forms:

> For concepts to exist, we must be aware that thoughts that involve them can be wrong, which in turn means that we must have a concept of thought. (Hinzen & Sheehan 2013: 58)
>
> What is missing in the mouse's thought is, we claim, not only concepts, but a formal ontology, by which we mean a system of formal distinctions by which

objects of reference are classified as objects and events, propositions and facts, properties or states. (Hinzen & Sheehan 2013: 46)

Quite clearly, from this perspective, propositional thought is an entirely linguistic business:

[…] the grammaticalization of the hominin brain in the evolutionary transition to our species uniquely explains why our cognitive mode involves a capacity for thought in a propositional format. (Hinzen 2014:226)

With grammar, we can refer and predicate, and the result is propositional truth. (Hinzen 2014: 235)

Not surprisingly then, the predicational format that Burge and Soames individuate, as we will see below, as the core feature of the representational power of *perception*, cannot be defined, according to Hinzen, in ways that are not linguistic:

There is also no known characterization of **predication** in non-grammatical terms. One could define predicates in formal-semantic or set-theoretical terms, but as Davidson (2005) discusses, the original grammatical predicates enter into the definition of the sets that then represent the predicates, which means that the definition explains nothing. (Hinzen & Sheehan 2013: 63)

These claims manifestly conflict with the results of mainstream perceptual psychology, on which some of the best theorizing in the philosophy of mind clearly depends (Burge 2005). The main result of this stream of research is that perceiving the world, on the part of a living organism, is producing *veridical representations* of states of affairs in the world (Burge 2010). The central notions in this view are the notion of *proximal stimulation*, for instance the registration of a certain distribution of light on the retina in visual perception, the notion of *formation laws*, that is, the computational principles that govern the transition from the registration of proximal stimulation to well-defined representational states, and the notion of *distal objects*, that is, the external physical sources of proximal stimulation. A central feature of this approach is that veridical perception depends not only on an explanation of how a given representational state is induced by a certain type of proximal registration (modulo the formation laws), but also depends, crucially, on the relations between proximal and distal objects, which are the subject of naturalistic inquiry.

Take the case of *vision* science. We know that the visual state in which an organism find itself depends on a complex set of formation rules, resulting

in a complex interplay of low-level, intermediate-level and high-level visual processing (Kandel et al. 2013). The retina transforms light patterns projected onto photoreceptors into neural signals that are conveyed through the optic nerve to specialized visual centers in the brain. In producing its output, the retina discards much of the stimulus information available at the receptor level and extracts certain low-level features of the visual field useful to the central visual system. Intermediate-level visual processing involves parsing the visual world into contours and surfaces that belong to objects and separating these elements from the background. This is a quite challenging task. When confronted with a complex visual scene, we might assemble the available features into a potentially enormous number of distinct objects. Nonetheless, we quickly classify these features into a set of objects that can be matched with internal representations of object shape and identity that are stored in the brain from earlier experiences.

High-level vision is also a complex matter. First of all, ascending visual pathways follow two parallel and hierarchically organized streams: the *ventral* and *dorsal* streams (Milner and Goodale 2006). In Chapter 3, we will see that this has important consequences for the way in which vision establishes reference to distal objects. For the time being, it suffices to say that object recognition is a complex task that can be decomposed into operations of visual feature integration into sensory representations of discrete entities and into operations of matching sensory representations of objects with knowledge of the object's meaning and function. These operations are encoded in different parts of the cortex, as confirmed by the observation that the lesions in the temporal lobe typically resulting, in primates, in the loss of the ability to recognize objects (*apperceptive* visual agnosia) have to be carefully distinguished from the lesions in the occipital cortical areas that are typically responsible for deficits that correlate with sensitivity to basic visual attributes such as color, motion and distance (*associative* visual agnosia).

All the complex computational processes by means of which certain retinotopic images are mapped into the visual perception of certain objects against a visual background are part of what we have labeled 'formation laws.' The error we should avoid is thinking that these formation rules are all there is, and that what matters, for perception, is only *proximal objects*. After all, we may think, being in a given perceptual state depends on the

physical states our brain is in and not on the properties of the distal objects that the relevant perceptual state is supposed to represent. A satisfactory account of perception – it may be added – must include the cases of perceptual illusion (including perceptual referential illusion) in which the only thing that matters, it seems, is the account of how a certain perceptual state originates from a certain proximal object, *modulo* the formation laws.

This view, primarily due to Descartes' influence, is seriously flawed. It corresponds to what we defined above, following Burge, 'individual representationalism.' The formation rules that are responsible for the mapping of a certain retinotopic image into a complex visual scene representing the ways in which certain distal particulars are endowed with certain specific attributes are not *private* properties of individual minds detached from the world, but rather the product of long-term phylogenetic processes that provide, generally speaking, a detailed evolutionary justification of the properties of these formation laws in terms of biological function. According to Burge and this important stream of research in the philosophy of mind, the main point is thus that

> In vision science, the idea is that when specific environmental conditions are realized and light from these conditions reaches relevant receptors in standard ways, where these ways are specifiable – mainly by laws of optics – and where certain specifiable proprioceptive conditions are met, the formation laws will, barring various kinds of interference, yield a perceptual state that specifies particulars as being in those environmental conditions. Such perceptual states constitute veridical perception – veridical seeing. (Burge 2010: 383)

What we have here is, in fact, an important argument for *methodological* externalism, according to which the mind is a part of the natural world and must be studied as such: sensory and perceptual systems developed in order to process and classify *distal* stimuli – objects and happenings in the world external to the mind. The objects of processing and classification are definitely *not* private inner sensory events. As we shall see below, there is nothing less Cartesian than this view of sensation as a complex outward-oriented computational process endowed with specific forms of representational power. From this perspective, the most important property of perceptual states is that they can be *veridical*, that is, they can offer a representation – however partial and 'perspectival' it may be – of what there is as it is.

As already hinted at, one of the most serious potential objections to this thesis is that perceptual states are the same, under the same set of law-like formation principles, both in cases of referentially successful perception and in cases of illusory reference (since the *proximal* cause of the perceptual state is somehow the same, quite independently of the existence of a *distal* cause – or so it may be argued). Once again, let us emphasize that this view is untenable. It is completely oblivious of the fact that formation laws are encoded in the brain of individuals as the result of the phylogenetic processes by means of which each species built up its own *representational* bridges to 'what is out there' through selection processes of independent physical variables that are certainly sensitive, in some way or another, to considerations of biological function within some specific ecological 'niche,' but that cannot avoid, by definition, to take the external physical world (or whatever a naturalistic style of inquiry delivers to us as *the* world) as the domain of the relevant distal stimuli. In Burge's words, 'objectified empirical representation precedes subjective representation both constitutively and phylogenetically' (Burge 2010: 402).

To go back to the 'snowflake fallacy' that constituted the point of departure of these considerations, the representational power of perception, as it manifests itself in specific cognitive events of attribution of properties to particulars, is completely independent of the capacity to distinguish reality from appearance, as well from the capacity to make those properties the object of higher-level representations in further cognitive events of property attribution.

Even more significantly,

> […] An important feature of empirical accounts of perception is that the general principles, laws, or operations determining transformations among informational and perceptual states need not be represented in the perceptual system (or by the individual) in any way (Burge 2010: 404).

This entails that perceptual systems can work perfectly without making their own operations, computational processes and laws of formation the content of further representational states, even 'implicitly.' What matters, in this view of perception, is the separation of the set of principles connecting *external* environmental conditions to *internal* proximal objects (like retinotopic images in vision) from the set of principles (formation laws)

that governs the transformation of the original proximal object into a fully-fledged representation of some aspect of the world.

If this picture is adopted, the 'snowflake fallacy' fades away. In fact, it is this picture of the *actual* workings of perception systems, stemming from mainstream research in perceptual psychology, that reveals the main source of the snowflake fallacy. To put it shortly, it is incorrect to claim that a languageless infant, in order to possess the concepts of 'flower' and of 'yellow' while performing (say, in vision) the attribution of yellow to a certain flower, should be aware of these concepts and be able to manipulate them in independent higher-order representations. This would be dangerously similar to proposing that the formation rules in perception should be conscious and constitute the object of – at least implicit – acts of cognitive/perceptual representation. This is clearly not the case. In patients suffering from apperceptive visual agnosia, the formation laws responsible for binding distinct visual features in the representation of single objects are arguably disrupted, but it would be completely ludicrous to suppose that these patients have an at least implicit representation of these laws and their disruption in their own mind. Physical laws operate in the world without any need to be represented. Similarly, formation laws operate in the brain without any need to be represented.

It is in many ways paradoxical to claim that concepts, in order to work properly within cognition, must be consciously accessible, as if this claim could be reconciled with a tradition of research in which it is strongly emphasized that *grammar*, in order to work, need not be consciously accessed. In fact, we think that Hinzen's claims can hardly be linked to Chomsky's overall view of language and cognition. This emerges quite clearly – we believe – from this quote from P. Pietroski, in which the Chomskian view according to which words activate a complex network of perceptual and abstract representations does clearly not conflict with the recognition that concepts are rooted in perception and are thus, as such, independent of words. In fact, to state that words should not be identified with the concepts they may be associated with is radically different from stating that concepts do not exist independently of words:

> Perhaps we could not think about (the various things that can count as) triangles, as opposed to merely being able to classify certain things as triangular, without two integrated and integrating capacities: an ability to lexically connect concepts

corresponding to perceptual prototypes, an abstract notion of space, and the idea of proof or necessity; and an ability to create sentential concepts unavailable without mediation by linguistic expressions that have the right features (Pietroski 2005: 273–74)

1.2 Language and the 'propositional' view of perception

Hinzen's claim that language is a necessary pre-requisite for the possession of concepts operating within propositional thought is based on the essential insight that percepts are pre-conceptual representations, and that they are such because:

a. They are not the result of intentional events of reference but of 'passive' elaboration processes;
b. They cannot be consciously accessed and become the source of further representations.

In the preceding pages, we have proposed that (b) does not constitute a valid argument for the claim that percepts are non-conceptual in nature (this was the argument against individual representationalism). In what follows we intend to show that:

i. Percepts, besides involving concepts, may give rise to 'propositional' representations;
ii. Percepts are not simply passive records of energy patterns that are incident upon the outer sensory receptors (what we might dub 'the Cartesian view' of sensation).

Starting with (i), the analysis of propositions recently developed by Scott Soames in a series of important contributions (King at al. 2014, Soames 2015), according to which propositions are cognitive events of ascription of attributes to particulars, explicitly endorses the view that cognitive acts of perception are essentially propositional in nature. More particularly, Soames argues that the classical view of *structured propositions*, traditionally identified with Russell's view of propositions, falls short of explaining why propositions should be held as inherently representational and, as such, bearers of truth, objects of propositional attitudes and, what is really the core issue, meanings of sentences. In fact, structured propositions are a special sort of logical representations for which the issue of meaning arises in the same way in which it arises for sentences,

conceived in terms of syntactic objects endowed with a set of well-defined formal properties:

> Just as the structural relations holding among syntactic constituents of a sentence show how they are to be understood, so the structural relations holding among the constituents of the proposition must show what it predicates of what. (King et al. 2014: 29)

If propositions are assumed to be what elucidates how certain formal structures can express a representational content, structured propositions will not do: what we need is an explanation of what these representational properties consist of, not simply an alternative (though maybe logically more perspicuous) formal structure endowed with representational power.

Similarly, the problem is not solved by the classical conception of propositions as functions from possible worlds to truth-values. According to Soames, this view will not do either, since it works only under the assumption that one already has an independent analysis of propositions and truth – that is, an independent account of what should be elucidated. Within the possible world approach, worlds are maximal states w the universe might be in (that is, each w qualifies as a property of the universe). Assigning a truth-value to w simply *correlates* to the proposition that predicates w of the universe and that is true or false depending on whether w holds or does not hold of the universe. The existence of a correlation does not provide a solid foundation for the concept of proposition; in other words, this account is completely circular:

> […] the possible-worlds conception of propositions fails to explain how propositions can be representational, and so have truth conditions, and […] it wrongly takes what it calls 'worlds' and 'truth values' as unexplained primitives from which it tries to construct properties and propositions, when in fact properties and propositions are needed to explain and illuminate both truth and worlds as world-states. (King et al. 2014: 35)

Soames' analysis is a truly radical departure from the classical Platonic view of propositions. It answers the question '*which kind of objects must propositions be in order to be able to count as the bearers of truth?*' by putting the perceiving and cognitive agents at the center of the picture. The answer is in fact that propositions are *classes of equivalence of individual cognitive acts of ascription of attributes to particulars*. In Soames' own words,

> Since the proposition that *o* is red is the event type in which an agent predicates redness of *o*, it represents *o* as being red because all conceivable instances of it are events in which an agent does so [...] Being inherently intentional, it can be the interpretation of sentences and utterances, without itself being the sort of thing for which an interpretation is needed. (Soames 2015: 97)

Given this view of what propositions are, at least in principle, perception will do as well as cognition or language. To illustrate this with an example, let us get back to our speechless infant. At the moment in which she builds up a visual representation of a certain flower as yellow, she ascribes the attribute yellow to that flower. It is this cognitive act of ascription, which puts the relation between the perceiving agent and a certain distal object at the center of the picture, that is veridical, and it is in this sense that propositions are bearers of truth.

The concept of proposition applies to individual cognitive events of property-ascription; and, in fact, the properties of a proposition (whether or not it is true) are the properties of those cognitive events themselves. For instance, the ascription of yellow to a certain flower within a cognitive event of visual perception is veridical if and only if the laws of optics have applied correctly and so have the laws of formation that govern the mapping from the proximal stimulus to the final perceptual state.

In this view, propositions can fulfill their role of truth-bearers because Platonic talk has been replaced with empirical talk and logical forms have been replaced by the well-formedness conditions governing the relation between proximal and distal objects and between proximal objects and final representational states. From this perspective, veridicality no longer holds of abstract objects (syntactic structures or logical forms), *it holds of individual cognitive events*.

We interpret Soames' view as entailing that the notion of 'proposition' becomes feasible (and successful) only when it is completely demystified: the matter is not which arcane logical structures may perform the function of truth-bearers (in which case the answer would be 'none'); the matter is rather which empirical conditions are responsible for the veridicality of particular classes of acts of perception and cognition. And in this case, the answer is: 'The conditions, investigated by perceptual psychology, under which proximal and distal objects are correctly related to each other and the laws of formation correctly apply.'

In a sense, propositions involve a shift from languages (sets of syntactic structures and logical forms) to concrete cognitive and perceptual events, where what matters is the relation between the perceiving or cognizing agent and the distal objects that, under normal conditions, bring about those events of perception or cognition (Fiorin and Delfitto 2020).

Propositions are thus truth-bearers (and successfully so) only because the truth-conditions associated to the sentence 'that flower is yellow' (essentially Tarskian biconditionals expressing that 'that flower is yellow' is true if and only if that flower is yellow) have been re-interpreted as the truth-conditions associated with the perceptual or cognitive event that the sentence 'that flower is yellow' is intended to express. These veridicality conditions are not a (Platonic) equivalent of the state of affairs in which that flower is yellow, but include the whole set of empirical conditions that must be satisfied in order for the entertained representation (in fact, the proposition) to be a *correct* representation of a certain state of affairs in the external world. As exemplified above, in the case of an infant's representation of a flower as yellow, these conditions contain at least the laws of optics and the laws of formation of the infant's visual state.

Demystifying propositions is mapping logic into perceptual psychology, so to speak. And this is tantamount to letting language (both syntactic structures and logical forms) get off its high horse. Sentences are not a higher-level repository of truth with respect to perception, they are just a way to express the truth, and *one among many*.

Put together, Burge's insights and Soames' insights provide us with a radically new perspective on language and perception, a perspective in which continuity replaces hierarchy. This new view is most conveniently exemplified by the contrast between Hinzen and Soames in their understanding of the notion of 'sentence.' In Hinzen's view, sentences are grammatical categories strongly detached from perception:

> Perceptual experience is richly structured, but does not reflect the categories of grammar, and 'sentence' is not an experiential category in any sensory-motor sense, similarly to an experience of pain, redness, or sadness. As we think, judge, and articulate, no sensory-motor experience of the world needs to be involved. (Hinzen & Sheehan 2013: 63)

For Soames, sentences are simply one of the possible vehicles for the expression of propositional thought, which is independently entertained:

> [Propositions] are pieces of information that represent things in the world as being certain ways; thus they have truth conditions. (Soames 2015: 95)
>
> [Sentences are] concrete events occurring at particular times and places in which agents produce auditory, visual, or tactile tokens endowed with semantic and syntactic properties [...] When S expresses p, one who understands S can entertain p by tokening S. (Soames 2015: 105)

Similarly, whereas Hinzen repeatedly emphasizes that predication is an essentially *grammatical* notion and its representational import can only be realized in language (as claimed in the quotes above), Soames is explicit about the fact that predication is the core property of *perceptual acts of cognition*. On this point, it is worth quoting Soames at some length, to get a more precise sense of his philosophical stand:

> Perception, like cognition, is, I think, a cognitive activity in which we do something that results in the world being represented in one way or another. Think of Wittgenstein's duck/rabbit example, or of an Escher drawing of a complex geometric structure. In the former case, a curved line can look either duck-shaped or rabbit-shaped. First we see it one way, then another. Once we realize that it can be seen both ways, we may try, with varying degrees of success, to move at will from a perception with one representational content to a perception with the other. The same is true when an Escher drawing of a building with a set of stairs that appears at one moment to be descending, and at another to be ascending. Similar experiences can be had in specially constructed rooms designed to create perceptual anomalies. To see what is before us first one way, and then the other, is to first predicate one property of what we see, and then to predicate a different property of it. Somehow perception makes properties and relations available to us to put together in different predicative patterns. (King et al. 2014: 95)

Ultimately, the view of perception we intend to subscribe to is the following. Perception is *representational*, hence *veridical* (Burge 2010). Perception, as a representation-building cognitive system, consists in cognitive acts of ascription of attributes to objects (Soames 2015). According to Soames, the types of these different cognitive events, which can take place both in perception proper and in cognition – say, both in vision and in language – are what is commonly referred to as *propositions*. For instance, the proposition that object o is red is the event type in which an agent predicates redness of o. Certainly, what one perceives is typically an object or an event; and still, the content of one's perception is a proposition (or a set of propositions, as in a complex visual scene) representing things as they are.

It follows that:

i. Perception is not only veridical, it is also propositional;
ii. Perception, to the extent that it can be characterized as involving the ascription of attributes to objects, is explicitly combinatorial and compositional.

There is more. We know, independently, that objects are not identified by binding one or more descriptive features at a certain location. Rather, objects are primitively 'tracked' (see the discussion of the notion of FINST in Chapter 3; Pylyshyn 2007, Fodor and Pylyshyn 2015), even in a certain number, at the same time, independently of the properties that they may satisfy. As for properties, they are identified as equivalence classes of *sensory* features (Burge 2010). This entails that properties do not simply reflect discriminatory skills. They rather give rise to mental signatures that respond to specific clusters of perceptual features (Reboul 2017). In this sense perception already involves *concepts*: the attributes that are ascribed to objects are, in fact, nothing else than *concepts* (on this see Fodor's review of Burge 2010, Fodor 2015). In this sense, from a Fodorian perspective, properties are not intensions – a collection of descriptive features each of which defines its own extension; rather, they correspond to the way in which the mind is singularly affected by the causal power of sets of sensory features that primitively define, for the purposes of perception/cognition, a certain extension.

An even more radical view, in which perception is hardly distinguished from sensation in terms of the complexity of the computational processes on which it is based, is put forward in recent work by M. Matthen (Matthen 2005; for a detailed discussion, see also Fiorin and Delfitto 2020). This view includes the so-called 'Classificatory Equivalence Thesis,' according to which the 'subjective feelings' that are proper to perceptual experiences are labels for physically grounded classificatory devices. This view explicitly contrasts with the traditional Cartesian thesis about sensation and higher operations of the mind. Here is how Matthen elaborates on the matter:

> Sensory consciousness is not a passive record of energy patterns that are incident upon the outer sensory receptors. According to the traditional Cartesian view, sensation lacks all attributive significance. It has no semantic value; it is not true or false, correct or incorrect. It is rather the material from which the epistemic faculties draw classifications or beliefs; the latter have semantic value, but not sensation itself. (Matthen 2005: 39)

> According to Descartes, animals possess neither the powers of conceptualization nor the conscious states of awareness that lie within the proper domain of souls. (Matthen 2005: 40)

It is sensation, and not higher-order cognitive systems such as language, that constitutes the epistemic source of objectification, conceived of, as in Burge, as the classification of objects or events in the external world – a view labeled as 'pluralistic realism':

> In one respect, the Sensory Classification Thesis is realist from the very outset: it insists that sensory systems classify distal stimuli—objects or happenings in the external world. Thus, it gives the senses an essentially outward-looking role. In this respect, it is sharply different from traditional Sense Datum Theory, which makes inner sensory events the objects of classification: according to this traditional perspective, red is a feature shared by the private and inner 'direct objects' of sensory events. (Matthen 2005: 19)

On these grounds, sensation itself is a complex computational process that feeds propositional and inferential knowledge:

> [Sensory awareness] supports induction, reidentification of external objects, and prediction. Our awareness of objects in three-dimensional space similarly affords us the ability to navigate the external world. These facts — the success of induction and of navigation —suggest that visual information must be anchored in real observer-independent features located in three-dimensional physical space.' […] The Sensory Classification Thesis claims, in effect, that sensory awareness can be expressed in terms of a set of singular propositions, messages to the effect that a particular individual is assigned to a certain class, and is identified as exemplifying a certain property. (Matthen 2005: 5)

And here is how Matthen argues that the perceptual grasp of a sensory-feature *F*, whereby an organism reacts to *F* in a certain way (innately or by learning), is clearly compositional in nature, based on the processes of feature-integration:

> A human's visual state when it attends to a blue disc is the result of 'binding' blue in a colour map to the co-located disc in a separate map of shapes. These two classifications are independent of one another. This implication of Feature Integration Theory accords with everyday experience: if one is able to entertain, in visual imagination for instance, images of a blue circle and of a red square, then one is able to entertain visual images of a blue square and a red circle. (Matthen 2005: 80)

Taken as a whole, Matthen's view of perception as a complex representational system holds that perception is endowed with highly abstract

computational properties, besides combinatoriality and compositionality, involving non-concreteness, altered topology, many-one mapping and loss of information:

> Neurons involved in coding a sensation (a red spot in the top-left quadrant of the retinal array) perform a classificatory activity (they respond to propositional conditions) and gather physically different conditions into psychologically relevant equivalence classes. The output of a neuron in the visual processing pathway indicates that its receptive field belongs (or fails to belong) to a certain class, the class of those that contain a certain retinal pattern, and suppresses additional information about its input. (Matthen 2005: 51)

This approach is empirically corroborated by the observation that different sensory systems are inherently sensitive to different features of distal objects. Monkeys, for instance, can be trained to respond to the larger of two rectangles, but not to the intermediate-sized one out of three rectangles. This in spite of the fact that the information corresponding to both of these conditions is present in the visual array. Arguably, the difference between the two conditions is that the first is discretely represented by the monkey's visual system and thus available as a key for conditioning, whereas this is not the case for the other condition (Dretske 1981: 151; Reboul 2017). According to Matthen, these observations are of course fully compatible with the realist stand on perception that has been introduced above, since in these conditions it is still 'natural and reasonable to say that some feature-detectors are responding to distal objects and events' (Matthen 2005: 59).

What is really striking, in Matthen's view of perception, is that the latter is not only propositional in nature, in a way that would probably be fully subscribed by Soames, but that it is already fully involved in the sort of epistemic relations canonically represented in propositional attitude structures. These complex relational structures, which Hinzen would argue are made possible by language, and only by language, are somehow already presupposed, as unconscious computational devices, in the structure of perceptual processing. More particularly, the activity of neurons can be characterized as assigning Boolean or Bayesian values to propositions. And since neurons' response conditions are propositions, visual processing can be functionally represented by propositional attitudes, and the transformations of the original conditions by inferential patterns (see also Pylyshyn 1986, ch. 6).

The 'internalist' nature of perceptual representations, as Chomsky would call it, is limited to the selection of those physical traits that are functional to the organism. The organism does not build up the world. Rather, it builds up a particular 'perspective' on the world. Still, what is represented *is* the world, *based on properties of the world and on properties of the mind*. In a sense, we might say, the story does not end up with mind-internal formation rules and proximal objects. The story crucially extends to distal objects, including *physical* variables as those revealed by scientific inquiry and *emergent* properties based on arrays of physical properties. Certainly, perceptual systems select the physical traits they are more 'interested in,' from an evolutionary perspective that emphasizes the role of functional adaptation to the environmental 'niche' in which the organisms express their perceptual abilities:

> As Kathleen Akins (1996) argues, sensory systems assess environmental events from a self-interested point of view: if a sensory system co-classifies two things, it is because these two things are comparable from the point of view of the organism they serve, not because anything else would so regard them, much less because nature would so regard them, whatever this might mean. (Matthen 2005: 60)

Because of this evolutionary perspective, which is in full agreement with Burge's sharp opposition to individual representationalism, Matthen's view of perception is eminently anti-Cartesian:

> To amend Descartes: there is something in the object, whatever it may turn out to be, which my visual system classifies together with other things that look blue (and which for this reason produces in me the sensation as of a blue thing). (Matthen 2005: 59)

From this perspective, the canonical example is provided by the functional definition of color vision:

> Colour vision is the visual discrimination capacity that relies on wavelength-discriminating sensors to ground differentially learned (or conditioned) responses to light differing in wavelength only. (Matthen 2005: 166)

There are points – we believe – in which Matthen's view on the propositionality of perception, though essentially correct, should be integrated and amended with Soames' view. For instance, the reason why sentences are simply the vehicle of propositions is not that propositions coincide with non-linguistic logical structures (this is the classical view, with which Soames successfully takes issue) but that propositions are the bridge, as we have seen above, to

the constitutive rules of perceptual psychology, which represent the objective conditions for the veridicality of specific classes of concrete cognitive events, in language as well as in perception:

> It seems reasonable to say that sentences exhibit syntactic structure, while propositions exhibit logical structure — after all, propositions are not linguistic entities, and syntax is linguistic structure. If this is accepted, one should be careful not to claim that only syntactic entities can be used to express propositions (Matthen 2005: 78)

1.3 Where we stand now: A look behind and a (puzzling) look forward

A number of conclusions clearly emerge from this discussion:

  i. There is no point in downplaying perception when assessing the role of language within cognition;
 ii. There is no point in regarding grammar as the cognitive source of objectification in representing the world;
iii. There is no point in endorsing a discontinuity thesis based on the hypothesis that combinatoriality, compositionality, propositionality, not to speak of representationalism, are grounded in language and are nonessential to perception;
 iv. There is no point in contending that propositions, as the repository of meaning and truth, are immediately identifiable with logical structures (set-theoretic constructions) and are as such primitively grounded in language; *foundationally*, propositions must be modeled as cognitive acts whose veridicality is primarily – even though not exhaustively – rooted in the correctness conditions of perceptual psychology;
  v. There is no point in contending that cognition is a snowflake, in the sense that it comes in as a whole – as is argued for syntax; having the concept of 'flower' does not entail having the concept of 'concept,' and entertaining a proposition in perception is perfectly possible though it does not entail possessing the whole set-theoretic apparatus that underlies linguistically expressed propositions.

These conclusions are incompatible with Hinzen's thesis that concepts are lexemes and that sentences are the necessary condition for the existence of propositions (and thought), rather than (merely) a formidable vehicle for the expression of propositions and for increasing the combinatorial potential of thought by means of a set-theoretic style of composition.

Perception and language have much in common. The interesting question to ask is, therefore, how much they have in common. A surprising finding, which will be presented in Chapter 3, is that some among the most abstract features of language, traditionally conceived, in fact, as a prerogative of the semantics of language, are already present in perception.

Does language mirror perception? And does this entail that even grammar is in continuity with the properties of perception? In the coming chapters, we will see that this is hardly the case, not only in virtue of the formal properties (discrete infinity, recursion, locality) that set syntax apart from perception, but especially in virtue of the abstract patterns that syntax arguably makes available for the expression of perceptual and cognitive content. Where do these patterns come from? We will argue that if there are some aspects for which grammar appears to mirror perception, there are also some aspects for which grammar appears to pre-encode a number of subtle epistemological distinctions between perception and cognition, which consciously emerge, as such, as the result of relatively advanced scientific inquiry.

What is grammar, then? Is it Alice's looking-glass or Alice's rabbit-hole?

Chapter 2   A philosophical interlude on perceptual content: Percepts, concepts, representations, and propositions

Before proceeding, we offer to the philosophically interested reader a short overview of the ongoing debate on the relationship between perception, belief, and knowledge. Without any pretension of exhaustivity, we will try to address the issues that are fundamental to the discussion to come, as they are presented and discussed in the current literature.

2.1 On the relation between perceptual experience and conceptual knowledge

What is the relation between the result of acts of perception (percepts) and what we usually regard as the primitive ingredients of thought (concepts)? If percepts have a conceptual substance, what is then the difference with respect to concepts and the process of conceptualization? If, conversely, percepts are devoid of conceptual substance, how can they feed propositional belief and justified belief, that is, knowledge? This was the main preoccupation of classical *foundationalism*: if knowledge has eventually to be anchored to our perceptual experience, and if the latter is no more than the subjective sense-data that we privately experience, how can then perception constitute a reliable epistemological foundation of belief and, ultimately, knowledge? More particularly, if perception is linked to the operation of the senses, and senses manage fundamentally analogical data, how can these data be put into an intelligible relation with the fundamentally digital operations of propositional thought?

Apparently, senses present contents that are continuously varying quantities – such as color, pitch, tones, timbres. Conversely, as Plato already observed in the *Philebus*, it is reason that presents definite measures; that this is the case is shown, for instance, by the *categorical* perception of phonemes with respect to the perceptual properties of speech (timbre, pitch, etc.). Something similar holds for colors. It is because of the *categorical*

jump between basic colors that the rainbow is perceived as banded, whereas it is actually a continuous wavelength gradient. Similar considerations hold for musical harmony. In a nutshell, this is the basic Cartesian received wisdom, shared by Hobbes, Locke and many others: the analogical realm of sense-data is incommensurable with respect to the digital domain of reason, and the general view adopted in this regard has been to suggest that *ideas* (reason) must provide a content to what is inherently contentless (sense-data, perception).

As we have already seen, there are many reasons to object to this Cartesian view of perception, and many of the positions we will briefly review below can be usefully elucidated by individuating them as pro- or anti-Cartesian.

All in all, we think it is fair to state that the philosophical positions that object to perceptual content are essentially disregarding (or underestimating) the empirical results of the most recent stream of empirical research in perceptual psychology, as T. Burge as forcefully emphasized. At the same time, there certainly are aspects of perception that cannot be immediately translated into propositional content (Dretske's 'simple seeing' is arguably a case in point), and we will have something to say about these cases in Chapter 4.

In this monograph, we argue in favor of the unity of cognition, by singling out features that are shared among perception, thought, language and conscious reflection about the operations of the human cognitive systems, that is, the theories of knowledge that we, as humans, consciously and effortfully build. But before further pursuing our endeavor, let us first take a look at the philosophical state-of-the-art.

An interesting overview of the issues involved in the question of what perception conveys to the perceiver is found in the miscellaneous volume edited by Berit Brogaard (Brogaard 2014). The general theme that emerges there from the different contributions is the relation of perception to cognition, emotions and action. Philosophically, as already hinted at above, the main topic is that of *perceptual content*. In perception the subject is arguably confronted with the proposition that the world is in a certain way: this is what could be defined as the *perceptual content* of perceptual experience.

From this perspective, perceptual content is a *representational state*. The question is then in which way this representational state interacts with or feeds *belief* and *knowledge* at the epistemic level.

This issue is strictly related to the issue of the *individuation of the senses* (Matthen 2015b). Under by now common neuropsychological guidelines, a *sense* can be defined as an information-gathering faculty that provides the subject with content-bearing experience that feeds belief, but, crucially, with the possibility that belief assent be withheld. To qualify as a *sense*, this information-gathering faculty must be subjected to a source condition – that is, to dedicated transducers selecting specific qualities of the impinging stimulus – as well as an 'integrated system criterion' – a system governing the interaction with information stemming from other perceptual and cognitive modules. A sense is thus a set of perceptual processes that begin from transducers specialized for extracting information from a particular kind of energy; *perceptual systems* can be defined, in turn, as systems of activities intended to bring about the same sensory experience. Within this model, the obvious criterion for understanding what senses have in common in different species is the biological notion of *homology* – the common evolutionary origin of pairs of structures in different *taxa*.

Those who admit that perceptual experiences have content and, henceforth, are inherently representational still have to spell out the identity conditions that apply to this content and their corresponding representations. According to some scholars, perceptual experiences are *propositional attitudes* (see for instance Glüer 2014). This claim is based in turn on the contention that perception is a *conscious* process and that acts of perception have a *conceptual* substance. Should we conclude from this that sensory processes must be conscious and conceptual in order to qualify as *perceptual acts*? Clearly, there are monitoring systems of sensory registration that directly feed motor systems and that are neither conscious nor conceptual in nature. A case in point is systems of homeostatic regulation, such as breathing. Another case is blindsight[4] and, more generally, the visual dorsal stream that directly feeds action, which is often set against the visual ventral stream that feeds belief systems in humans (Milner and Goodale 2006).

---

4   This is the capacity of cortically blind subjects to respond, to some extent, to visual stimuli which they are not able to verbally report, that is, which they are unable to consciously experience (Weiskrantz 1986).

Notoriously, the issue of consciousness is a delicate one. Consider for instance the bee visual system. Bees probably do not have *access-consciousness* (in the sense of Block 1995): there is no way for them to report, either in thought or in some language-like medium, about their sensory-perceptual states. However, bees are probably endowed with some sort of *phenomenal* consciousness: to see and to be attracted towards a flower *feels like something* (to use Nagel's famous expression) to them, though we do not have any idea which kind of *qualia* are actually involved. On a similar vein, it would be wrong to assume that a bee's visual system directly feeds the bee's motor system, in the way bacteria are mechanically attracted to de-oxygenated areas through magnetic field sensory registration. In fact, there are reasons to believe that bees are actually endowed with a working space where the representational content provided to them by their visual system may interact with the independent information provided by other sensory systems. From these remarks, we might conclude that the bee's visual system might qualify as a fully-fledged perceptual system, though it is not accessible to consciousness (again, in Block's sense) in the way considerable parts of the human perceptual systems undoubtedly are.

The content view of perception is the view that we explicitly adhere to and support in this monograph. However, we should not underestimate the fact that the competing view, according to which perception has *no content*, also has many representatives. If representationalism faces the challenge of explaining how perceptual content exactly relates to belief and knowledge, anti-representationalism – the view that perception has no content – also faces many challenges, one of which is represented by the *multi-modality of perception*. If one sees a flash and hears two beeps, then she will see two flashes: this appears to entail that two representations from two different sense modalities are matched at some early phase of perception; and if there is no perceptual representation as such, it is difficult to establish what it is that is matched.

The debate between representationalism and anti-representationalism hinges crucially on which features of the environment are informationally processed in perception. Certainly, what is processed is relatively low-level features of the environment like color, texture, smell, etc. According to Siegel (Siegel 2014), affordances – those features of the environment that solicit action from the part of the perceiver – are also part of the

representational content of perception states. In fact, *enactivism* holds that perceiving action is tied up with performing it. For instance, J.J. Gibson claims that we perceive affordances, possibilities of actions that objects afford to us (a view related to Husserl's notion of *life-world*). The issue is whether this sort of action-driven information processing is also directly or indirectly linked to a representational format.

Interestingly, Siegel (Siegel 2010) advocates a strong representational view of *perceptual experience*. According to Siegel, *visual phenomenology* is rich, and this thesis provides the foundations for the so-called 'Rich Content View,' according to which vision does not only represent low-level features such as color, shape, illumination and facing surfaces but also higher-level features such as personal identity, causation and kinds of objects. Phenomenal conscious states include, besides sensory states, bodily sensations, moods, visual imagery, cognitive experiences. Among these, there is an inherent disposition, in phenomenology, to single out kind-level properties, as when we recognize, in reading, known characters with respect to characters belonging to an unknown writing system, or when we recognize, in visual experience, apple trees with respect to trees that we cannot appropriately categorize. In other words, visual phenomenology brings us directly in touch with important forms of higher-level conceptualization. Importantly, this entails that perceptual *experiences*, being conceptual and, at least in principle, *propositional*, are also, and crucially, endowed with a semantic value, that is, they are *veridical* (accurate or inaccurate *semantically*).

We may propose, in fact, that the content of a perceptual experience consists in the ascription of a property to an object – that is – presents a certain cluster of properties as being instantiated. In visual acts of perception, this is done by binding the feature maps that correspond to properties to primitively tracked objects (Fiorin and Delfitto 2020: Chapter 33 and the references cited therein). The possibility arises of characterizing some sorts of visual experiences as *de dicto*, as when the perceiver cannot establish how a certain cluster of properties is instantiated, though he is conscious of the fact that it is instantiated. A case in point is Dretske's case of the moving train, where we sit looking out of one of the windows and we cannot tell which one of the two trains is moving, the one we sit in or the one we are seeing through the window, that is, we cannot tell *how* the relevant

movement property is *individually* instantiated. This leads to the conclusion that visual phenomenology is not only conceptually rich and characterized by a propositional format, it also anticipates certain high-level semantic features of language, such as the *de re/de dicto* dichotomy, at least if we roughly identify *de dicto* with the access to some descriptive content that is not obviously instantiated in a publicly recognizable individual object.

As for the ontological status of phenomenal states, the fact that a phenomenal experience in which a cluster of properties is instantiated (as in veridical instances of correct perception) be phenomenally indistinguishable from a phenomenal experience in which the same cluster of properties is not instantiated (as in illusions or hallucinations) leads Siegel to favor *internalism*. This entails that for Siegel *disjunctivism* – the position that defines perceptual experiences as relations between the perceiver and external objects in the environment – cannot be correct in proposing that strongly veridical experiences (instances of correct perception) and weakly veridical experiences (illusions or hallucinations) are ontologically distinct *as phenomenal states*.

Notwithstanding, we believe that Siegel's position on the ontological status of perceptual experiences should not be assimilated to the rejection of the view, endorsed among others by Burge and Matthen, according to which perception constitutes, epistemically, the primary source of *objectivity*. The fact that a hallucination is phenomenally indistinguishable from a strongly veridical act of seeing is still compatible with the thesis that the particular phenomenal state exemplified by the hallucination phylogenetically correlates with a state of affairs whereby certain distal objects have certain properties. In other words, even purely internal phenomenal states have an 'externalist' justification in phylogenetic terms, based on their correlation and similarity with externally motivated rules of formation of perceptual experiences, as they are defined by perceptual psychology. At the same time, the *inductive* force of the formation laws and principles governing correct perception explains why veridicality is only probable and never granted.

How a potentially dichotomic characterization of the phenomenology of perception can be made compatible with the 'externalist' view that perception provides veridical representations of the external world is shown by Cohen's analysis of some of the cases in which perceptual states provide the perceiver with conflicting information (Cohen 2015b). This is what

happens, for instance, when we perceive a uniformly painted but not uniformly illuminated wall: we can distinguish the color appearance of different regions of the wall, while perceiving the whole wall as having the same color. Even more compellingly, when we perceive the perceptually contradictory Escher drawings, we attribute incompatible properties to one and the same object; when, in contrast, we perceive a tilted penny in water, we attribute compatible properties to the penny (it is tilted *and though* not tilted). This puzzle of contradictory information in perception is effectively summarized by Peacocke as follows:

> Your experience represents these objects [trees] as being of the same physical height and other dimensions; that is, taking your experience at face value you would judge that the trees are roughly the same physical size […]. Yet there is also some sense in which the nearer tree occupies more of your visual field than the more distant tree. This is as much a feature of your experience itself as its representing the trees as being the same height […]. (Peacocke 1983: 12)

Is this dichotomic characterization of perception a reason to adopt a strong internalist stand according to which the properties of perception do not reflect the properties of external wordly items? There is of course a sense in which the perceptual state is obviously a property of the perceptual system, an 'internal' property of the mind, as opposed to distal, worldly items. Cohen's position is a *relationalist/contextualist* view according to which distal, worldly items that we perceive have dispositions to cause those perceptual states in us under certain circumstances; and these dispositional properties — henceforth, perceptual state dispositions — are properties of distal, worldly items as opposed to perceptual systems. As a consequence, there is no inconsistency in a thing being simultaneously round and disposed to generate in us the kind of perceptual state we undergo when perceiving an ellipse straight on. In this way, what we have in perception is both 'public' object features and 'perspectival' features: what should be taken into consideration is the 'objective' disposition objects have to generate certain perceptual states under well-defined conditions.

The kind of 'color relationalism' defended by Cohen is in the same spirit (Cohen 2006). This *relational view of colors* holds it that there is a range of variation in the chromatic effects of perceivers that all satisfy the standard psychophysical criteria for normal color vision. Veridicality is thus ensured within this range. The general issue is perceptual variation: when

the color perception is different depending on the surroundings, there is solid ground to discard one of the two perceptual representations as non-veridical. Cohen's approach has the important advantage that ontological preoccupations are countered by recourse to *semantic contextualism*: The predicate 'is yellow' is relativized to a number of context parameters. Colors are defined as functional roles connecting distal stimuli to the visual system: x looks red/green etc. to subject S in C if and only if, by visually attending to x in C, S is caused in C to have an experience of red/green etc. In this way, perception emerges as the disposition of *external* items to manifest themselves depending on the contextual features regulating their interaction with the internal properties of the perceiver. Perceptual states are properties of the mind, but they are also, and crucially, a function of the independent properties of external things, to be objectively defined in terms of *dispositional/relational features* these things are endowed with.

Intermediate positions between representationalist and anti-representationalist views of perceptual content are also possible and in fact largely attested. One of these is the *relational* view of perception. One of its recent advocates is Schellenberg (Schellenberg 2014), who contends that perceptual experience need not have representational content, though it *may* be associated with representational content, in that sentences/propositions may be used to describe perceptual experiences (though these propositions are not part of the perceptual experiences themselves).

As already mentioned above, *disjunctivism* is the view that perception relates perceptual experience to external objects in the environment. This view, also referred to as naïve realism, is based on the notion that acts of perception include the referent or distal object as one of the basic components of the relation, to the effect that hallucinations are (by definition) disjunctive with respect to veridical acts of perceptions. However, one familiar argument for anti-representationalism comes from the observation that hallucinatory experiences are of the same sort as correct events of perception, as Siegel correctly emphasizes. Do hallucinatory states have content? As the reader may suspect, there is a large variety of answers to this question (see Chapter 3). According to Tye, for instance, they have token-content but lack type-content (Tye 2014). According to Matthen, the content of perceptual experience is *imagistic* and has the propositional form of an existentially quantified proposition expressing that there is an

object *o* of sortal type S that instantiates feature F and occupies position L (Matthen 2014). In our opinion, the most convincing line of analysis is in this respect that advocated by Siegel and Cohen, in which what we consider to be rather obvious externalist requirements are elegantly accounted for, though these two authors reach in fact distinct ontological conclusions (Siegel 2010, Cohen 2015a,b).

We should also mention authors according to whom *perceptual content* is distinguished from *cognitive content* due to the phenomenal intentionality that characterizes the former – the kind of *aboutness* that is associated with perception and is inherently grounded in the subjective, experiential feature of certain mental states. A case in point is the position held by Terry Horgan, according to whom brains in a vat could not perceive color but could think about it (Horgan, Tienson and Graham 2004). And there are also authors who believe that perceptual content is not preserved across different sense modalities (Budek and Farkas 2014).

2.2 Cognitive penetration

Those who wish to defend the view that visual perception has representational content must face the following fundamental question: What does visual perception exactly *represent*? From low-level properties (such as motion, shape, color) to high-level properties (such as faces, values, affordances), to limit ourselves to standard ontological vocabulary. According to William G. Lycan, the demarcation line between perception and cognition is not clear, and the real conundrum is the issue of *cognitive penetration*: To which extent are the non-purely cognitive interpretation features of perception the result of cognitive penetration?

A number of philosophers concede there is no guarantee that the content of a perceptual experience is the same as the content of the *judgment* based on that experience: we need to understand the experiential condition of the perceiving subject as it stands prior to the subject going on to judge, on the basis of experience, this or that about how things stand in the subject's perceived environment. As emphasized by W.G. Lycan, we should acknowledge the possibility that, in exploring perceptual experience, 'attention alters the first-order state, or that it adds content' (Lycan 2014). The traditional demarcation line between perception and cognition seems thus related to

the broader concern that perception apart from description lacks content and it is only our attempts to *explain* it which lead us to think otherwise. It is when we describe perceptual experience (in language or thought) that we give it a content.

This 'conservative' view of perception is probably rooted in the classical view of perception as an inherently incomplete process that needs cognitive complementation. According to the traditional *Receptoral Image Model* (RIM) of perception (Matthen 2015a/b), sensory information, that is, activation of the receptors on the retina, is assumed to be equal with the 'perceptual given.' The conceptual ingredients of perception are provided by perception-external cognitive capacities: depth, for instance, is provided by independent inferential devices or learned association (since visual perception is essentially two-dimensional).

This traditional view of perception is slowly fading away but is still persistent in common thinking and common talk. It probably borrows its original force from Sellars' criticism (Sellars 1956) of the 'myth of the given' – the *foundationalist position* according to which first-personal perceptual experience primitively feeds epistemic belief. Sellars contends that basic beliefs expressed in the form of subjective experience reports of the form of 'That looks red to me' need be put in the conceptual space of inferential reasons in order to be able to provide an adequate foundation to perceptual knowledge. Perceptual knowledge is thus cognitively-based knowledge and non-inferentially justified 'basic beliefs' are untenable.

More generally, we should acknowledge that the claim according to which certain higher-level aspects of perception, such as depth, are fallibly inferred, together with the claim that hallucination is subjectively indistinguishable from ordinary perceptual experience, typically led to skeptical positions about perceptual knowledge.

These skeptical positions are easily overcome if one adopts the view that the higher-level components of perception are in fact not external to it, and more particularly the view that intermodal and cross-temporal integration are not extra-perceptual mental operations. Matthen 2015b persuasively shows that this view is supported by the result of neuroscientific investigation of speech and music perception. Experimental results show further that not only are there specialized modules for low-level features, such as color and shape: in fact, Gestalt psychologists had already demonstrated

that there is a perceptual basis for causality connections (contrary's to Hume's stand), as in contiguity and movement relations among abstract geometrical objects; and that, even more generally, the perception of movement and objects is not based on perceiving all the temporal and spatial parts of these entities. Moreover, single-cell neuronal recordings and the application of fMRI techniques show that that there are also specialized neuroanatomical areas for the extraction of content relevant to identification of specific kinds of higher-level objects. High-level categorization is thus internal to the perception process. Seeing faces, hearing melodies, 'inferring' causality can be shifted from the domains of learned association and rational inference to the domain of perception itself. On the one hand, *visual agnosia's* clearly demonstrate that perceptual deficits can show up even in the absence of receptoral defects. In particular, *apperceptive agnosia* manifests the failure of object recognition not only when the basic visual functions (color, shape and motion) are intact but also when, crucially, language and higher-cognition skills are normal. On the other hand, *multimodal perception* and multisensory integration, as in the rubber hand illusion (Botvinick and Cohen 1998) and in the illusions created by combining intersecting geometrical shapes with auditory stimuli (Cohen 2015b), make it difficult to explain how early matching between different sense modalities could take place without the support of a perception-internal level of propositional representation. To exemplify, as already hinted at above, if one shifts from seeing one flash to seeing two fleshes when she hears two beeps, it is arguably because cross-modal perceptual representation starts very early in the perception process.

In this way, the skeptic notion of the 'veil of ideas' is overcome. That is to say, *concepts* are not barriers between sensory impressions (subjectively experienced sense-data) and the external world, they are, so to speak, part and parcel of the constitution of perception.

True, *vision* is based on light frequency and light reflectance; notwithstanding, it targets complex objects and their spatial distribution, as well as complex events and the relationship between those objects and events. Similarly, *audition* is hardly definable in absence of higher-order conceptualization: it is certainly based on low-level features such as sound frequency; notwithstanding, it targets the temporal order between events and the properties of the objects that participate in those events. *Touch* targets the

abstract notion of 'pressure,' since it is through pressure that the lower-level features of texture, weight, hardness, etc. are manifested. And perhaps even more convincingly, *taste* is the complex product of the integration of the basic taste receptors on the tongue and olfactory skills (Matthen 2015b).

In these conditions, the traditional image, in Western philosophy, of senses as passive receptors in need of conceptual complementation is difficult to hold. It rather seems that concepts – and the concomitant propositional format that goes with conceptual representations – are in fact the essential ingredient of perception.

If this does not entail, in itself, that perceptual experience needs to be endowed with *conscious conceptual content*, it certainly supports the view that perceptual processes are inherently representational – pretty much in T. Burge's sense (Burge 2010) -, in that they are not only susceptible of being described in propositional terms, but in the stronger sense that a conceptual/propositional format is constitutive of the senses as specific information-gathering faculties.

To this, the informed reader might counter – associating with the anti-representationalist stream introduced above – that philosophers like Wright (Wright 2015) have strongly contended that all these high-level properties are necessarily the result of *cognitive penetration*, exactly the issue under discussion in the present section. In fact, a whole stream of neuroscientific research on vision informs us that *early vision* is characterized from representations that are impenetrable to conscious report in language and/or thought (in Block's sense of 'access-consciousness'; Block 1995). From this perspective, *late vision* would be conceptual only because there is *penetration from cognition*, in terms of top-down processes (in neural terms, from the frontal cortex to the inferior temporal cortex).

However, Matthen convincingly argues (see also Fiorin & Delfitto 2020) that proximal stimulation is filtered by means of inherently propositional classificatory devices, associated with specialized neuronal circuits. So, a denial that early vision is conceptual is unwarranted. It reduces to the claim that the categories involved in early vision are not conceptual just because they cannot be accessed by conscious thought. As we have seen, these claims are unwarranted since they constitute the main ingredients of what we have dubbed (following Burge) 'individual representationalism.' In fact, if early vision is characterized by unconscious patterns of propositional

classification, cognitive penetration can no longer be defined as the complementation of passive patterns of sensory registration with higher-level cognitive properties; rather, cognitive penetration would reduce to the observation that low and high cognition are actually indistinguishable.

True, the categories involved in early vision are not open to consciousness, in Block's sense that the classificatory devices to which they give rise cannot be reported in language or thought (see Noë and O'Regan 2002 for a neurocognitive perspective). Notwithstanding, this cannot be an argument in favor of the non-conceptual status of these very same classificatory devices. It should simply be taken to support Burge's view that the laws of formation in perception are not themselves part of the perceptual representations, a point that seems to have a general validity across cognition, and that is highly reminiscent of Chomsky's notions of knowledge of grammar and rule-following.

A different potential criticism concerns the observation that there are aspects of perception which are non-conceptual. A case in point is Pylyshyn's process of *object-tracking*, which is not property-driven, hence arguably devoid of conceptual content. We believe that this point is moot, since visual indexing in Pylyshyn's sense represents a primitive, non-descriptive modality of concept-formation, within the more general view that demonstration is a primitive modality of referring that independently feeds propositional representations.

In this respect, a similar line of analysis is developed by Dickie in his reflections on demonstrative thoughts (Dickie 2015). According to the traditional view, object-identification is not delivered by (visual) perception, but by the action of higher-cognition on the unstructured output of the primary visual system. In this way, the traditional view clearly favors the *descriptive view* of demonstrative thoughts with respect to the alternative *acquaintance view* (in the latter, demonstrative reference abstracts away from descriptive content). Dickie argues for the so-called post-traditional view, according to which objects are singled out in a way that abstracts away from the properties they are endowed with. More particularly, perception directly delivers the sort of three-dimensional objects about which demonstrative thoughts revolve. A belief of the sort 'that is F' is justified by perception if and only if perception correctly links the perceiver to the object F is about.

Similar considerations hold for compositionality, which should also fall within the scope of perception. In particular, Evans' argument that perception is not conceptual because conceptual representations are inherently compositional (Evans 1982) does not go through, since it is effectively countered by Matthen's arguments that perceptual content is demonstrably compositional (Matthen 2005). To a certain extent, this is already revealed by animal perception: for instance, a bird that has been trained to peck at blue squares also pecks at a blue disc.

The conclusion we should draw is that the view according to which *reason* and *perception* are radically different modules of the mind is completely unsupported. We could call it the Grand Illusion of Modern Western philosophy. It was Descartes who held the view that what is representational is the *judgment* that we draw from the sensations occasioned on our sense organs by physical motion, a view permeated by his negative attitude towards perception as a passive device of sensory registration. As observed by Matthen, Plato and Aristotle had in fact advocated the opposite view that perception is essentially representational. Aristotle justified the representational view of perception in terms of a transferal of form from the distal object to the mind, a perhaps too 'externalist' position in modern terms, somehow reminiscent of Berkeley's theory of vision according to which the retinal image is built by exploiting associations with the distal objects. In spite of these limitations, both Plato and Aristotle held the view that perception is representational and propositional, though they disagreed on the reliability of the delivered propositions: for Plato, perception and reason compete, for Aristotle – they integrate. In this monograph, we intend to challenge the traditional dichotomy between low and high cognition. We will adopt the Aristotelian view that perception and reason integrate, under the radical insight that the formation laws of perception, though unconscious for the perceiver, are actually pervaded by many of the features that we would be inclined to reserve for reason.

2.3 Modularity

The approach we have just outlined raises subtle issues concerning *modularity*. An interesting discussion about the demarcation line between low- and high cognition is found in Deroy (2015). According to the received

wisdom, low-level cognition is characterized by shallow outputs and is *strongly modular*, in the familiar sense that it tends to involve automaticity, encapsulation, unconsciousness, and all other typical features of cognitive modularity (Fodor 1983). Visual systems, for instance, and most notably early vision, would be characterized by the presence of a primitive, pre-semantic, non-interpreted level of representation. Deroy convincingly argues against this view. Granted, color and shape are processed separately and unconsciously in early vision, by means of dedicated and distinct neural circuits in the visual cortex. However, it is also true that they are bound together already at the level of low-level visual processing, frequently by means of intra-modular holistic and top-down effects, as when different shades of color assigned to a certain object (say, a cup) are experienced as the *same* color (and, crucially, not just because they belong to a single object). Another case in point is the so-called 'light from above' prior: the same figure is seen as concave or convex depending on the non-cognitive intra-modular assumption that light comes from above. Quite similar intra-modular holistic traits play a crucial role in *speech perception* and *face perception*. Further challenges to modularity are provided by synesthesia, by the traditional notion of cognitive penetration – the influence of propositional thinking on perception, which has been discussed in the preceding section – and by multisensory perception.

   All this evidence suggests that modularity can be rescued only at the price of reducing it to the notion of *nano-modularity*, based on a large number of interacting micro-modules. More generally, biological models of functional explanation are replacing the Fodorian digital models of the computational mind; within these developments, a central role is played by *embodied accounts* in which high-level mental features such as concepts and categories are no longer assumed to be orthogonal to the perception and motor systems, as part of the complex of bodily interactions with the environment. Finally, it should be emphasized that more or less definitory features of modularity, in the classical Fodorian conception, are actually ill-placed. This might be the case for unconsciousness: Fodor himself indeed used the permanence of *conscious visual illusions*, like Muller-Lyer,[5] to convincingly

---

5   This is the optical illusion (first created by the German psychologist Franz Carl Muller-Lyer in 1889) in which two lines of the same length appear to be of a

argue that visual perception of size is modular, and thus impermeable, as such, to cognitive penetration.

Clearly, evidence is rapidly accumulating in favor of an intramodal and intermodal integration of sensory information. This requires a new integrative view of perception, as also confirmed by linguistic effects such as the so-called *McGurk effect*.[6] Cohen (Cohen 2015a) interestingly argues for a new definition of modular encapsulation, based on the idea that a process is modular if it is *anisotropic* – that is, if the range of its input parameters is delimitable, in the sense of being defined by a finite array of basic ingredients. For instance, illusions based on the perception of apparent motion are modular, since they are based on a process reducing to the integration of a finite array of basic elements: contrast, luminance, and object information. The same holds, in fact, for the Muller-Lyer illusion, whose input are lines in a certain well-defined arrangement. Conversely, the systems of prospective memory and action control are non-modular (i.e. they are *isotropic*), since they are open to all kinds of perceptual information. Similarly, analogical processes are obviously non-modular. From Cohen's perspective, the most interesting case is perhaps the case of categorization, which is generally believed to be perception-external. However, the case of 'chasing' (in which subjects categorize a visual array of purely geometrical figures as 'chasing') shows that categorization might promptly qualify as perception-internal. At the same time, Cohen also suggests that there are cases where higher-order classification is *anisotropic*.

All in all, it seems fair to conclude, at least as a methodological guideline for further research, that (i) Fodorian modularity is undergoing extensive revision and that (ii) new versions of cognitive modularity do not constitute

---

different length, essentially depending on whether the arrow fins are protruding outwards (in which case lines appear longer) or on whether they are pointing inwards (in which case arrows appear shorter).

6   This is a powerful multisensory illusion in which a voice articulating a consonant is dubbed with a face articulating another consonant. The illusion consists in the fact that even though the speech signal is well recognized in isolation, when dubbing takes place it is heard as a third consonant, different from both the consonant corresponding to the acoustic stimulus and from the consonant corresponding to the visual stimulus. This effect was first reported in McGurk and MacDonald (1979).

a serious obstacle to integrative views of low- and high-cognition. To the contrary, these new versions of modularity seem to presuppose that there is no stable demarcation line between perception and reason – to use the classical philosophical terms.

2.4 Perception, concepts, and categorization

Something more should be said at this point about the relationship between perception and *categorization*. Endorsing a content view of perception entails assuming that percepts are endowed with conceptual content. But the relevant question is, of course: How is this the case exactly?

Notoriously, this is an impervious domain, and the problems actually already start with the very notion of *concept*. Concepts have been regarded as prototypes (Rosch 1973 and subsequent work), as exemplars (Brooks 1978; Medin and Schaffer 1978) or as theories (Carey 2009; Murphy and Medin 1985; Gopnik and Schultz 2007), up to the eliminativist approach defended in Machery (2009). All these approaches face serious difficulties, whose primary root is their inability to effectively meet the obvious stability and compositionality requirements that arise when conceptual representations are identified with the way meaning comes to be expressed.[7] Two classical alternatives are Millikan's view that assimilates concept endowment with the ability to categorize the objects that fall in the concept's extension (Millikan 2000) and Fodor *radical atomism*, according to which concepts are mental signatures that enter mental digital computations (Fodor's *Language of Thought*), whose meaning reduces to the *extensions* to which they correspond (Fodor 1975).

Millikan's assimilation of concepts to *categories* or, better to say, to the capacity of categorizing on the part of the perceiver or cognitive agent, is in many ways palatable, especially for those who adhere to the view that the roots of categorization lie in perception. In fact, human perceptual abilities do not reduce to visual discrimination but massively proceed to categorization: to cut a long story short, humans are able to scrutinize complex visual scenes and to induce complex concept hierarchies on them, proceeding from

---

7   The classification of theories of concepts adopted here is largely based on the discussion in Reboul (2017).

the more abstract (for instance, general shape configurations that abstract away from a rich set of fine-grained details) to the more concrete (the fine-grained set of features corresponding to less general categories). More particularly, a by now increasing series of comparative studies on human/animal perception reveal that humans have a superior ability to employ, in selection and discrimination tasks, categorizing tools that tend to fall within Rosch's notion of 'basic-category,' whereby objects are judged as belonging to the same category if they share, roughly, the same motor-pattern and the same general shape. For instance, whereas primates other than humans tend to group different breeds of dogs under different categories, they may have (unsurprisingly, with the partial exception of chimpanzees) problems to group them under the basic-category 'dog' (Reboul 2017 and the references cited therein).

So, why do not assimilate concepts to categories and to categorizing, as Millikan does? The problem with this strategy is that it's quite easy to find cases of concept possession in which the latter does not translate at all in the ability to recognize the objects falling under the concept's extension as belonging to the relevant category. I might be said to possess the concept of 'beech tree' (I can use it in language and thought), but I might not be able to decide which trees are in fact beech trees. Similarly, I might be said to possess the concept of 'electron' (I can make plausibly correct assertions concerning the role of electrons in certain models of the atom) but of course this does not easily translate in my capacity to (perceptually or cognitively) discriminate electrons from other subatomic entities.

Fodor's atomism is also problematic. In particular, his strategy faces fundamental difficulties as it rests on the assumption that 'semantic relations between the mind and the world are 'grounded in' causal relations between the mind and things in the mind's perceptual circle' (Fodor and Pylyshyn 2015; p. 126). For Fodor, conceptual content boils down to mental representations (metaphysically, all there is in a concept is its status as mental signature/representation), under the crucial assumption that the reference of these mental representations supervenes on the causal relations with the objects the mind is exposed to within 'the perceptual circle.' This has been difficult to digest to many. In a sense, Fodor does with concepts what D.M. Armstrong did with perceptual knowledge (as we will discuss below): as there must be a causal relation between unconscious perceptual

knowledge and conscious epistemic knowledge (Armstrong 1973), so there must be a causal relation between things in the perceptual circle and the concepts we develop by being perceptually exposed to those things.

The problem has been – and surely partly still is – that we would like to inquire into the nature of these relations, in order to give them a serious empirical content. A recent stream of research in perceptual psychology is doing exactly that, though virtually everyone would agree that there still is a long way to go. For humans, visually inspecting a complex visual scene consists in inducing from it a complex propositional pattern whereby an event of a certain kind involves the participation of objects of a certain kind: which concepts instantiate this propositional format depends on the application of a complex set of unconscious perceptual and cognitive mechanisms, of which endogenous attention constitutes part and parcel. Or take, more compellingly, the concept of 'object.' Fodor and Pylyshyn 2015 empirically substantiate the empirical thesis (based on Pylyshyn's long-standing scientific results) that what we call objects (better to say, the objects that fall under the concept 'object') are just 'things that can be tracked,' whereby *tracking* is an undefined *perceptual primitive*. The notions of 'object' and 'tracking' are in this way '*causal mind to world processes,'* and conceptualization – intensions, descriptions, and all the rich epistemological apparatus traditionally associated to concepts – can be completely dispensed with.

2.5 Perception, reference, and language

Whatever the reader might think of this endeavor of explaining away concepts through the findings of empirical psychology rather than through the full array of assumptions of *a priori* epistemology, there is a point here that deserves in our opinion to be further developed. In Fodor's view, *reference* has no meaningful content (since for him all there is to meaning is reference), in the sense that the relation between things and concepts or words must be freed from the unbearable epistemic and cognitive burden of the traditional descriptive apparatus of definitions, intensions, and so on; to the contrary, reference should be inquired as a subject matter of empirical psychology, and more particularly of perceptual psychology (that is exactly what Fodor and Pylyshyn 2015 claim they are doing). The point to which we would like to draw the reader's attention is that there

might be ways to generalize this perceptual approach from concepts to *meaning as a whole*. As a matter of fact, sentence meaning (and even more, Grice's notion of 'speaker-meaning') is commonly regarded as the result of sophisticated conscious cognitive strategies. The speaker of a language is assumed to be endowed with a conscious awareness of 'what-is said' by using a sentence in a language. Moreover, under usual Gricean assumptions, 'what-is-said' is consciously enriched with a number of variable contextual elements, a significant part of which crucially involve the speaker's communicative intentions and the hearer's disposition to determine what these intentions are, by means of the conscious deciphering of gestures, tone, etc. of the participants in the relevant acts of communication. To this traditional Gricean picture, Azzouni opposes a *phenomenological analysis* of sentence-meaning that develops a clear-cut *perceptual* notion of what sentence meaning actually is (Azzouni 2013). According to Azzouni, both *what is said* and *what is implicated* depend on the inclusion of contextual variables. However, *phenomenologically* what is said corresponds in fact to viewing an object from a certain perspective: this is done automatically, by an *unconscious* detection of the contextual variables. To the contrary, what is implicated entails that there was a meaning shift brought about by the conscious detection of the relevant contextual variables (such as tone, gestures, or even the incompatibility of the word meaning with the experienced situation).

Azzouni does not see, while opposing the neo-Gricean mainstream, the ways in which contextual elements infiltrate what is said as due to the speaker's intentions. Rather, what is said by an expression is something the speaker-hearer *involuntarily and spontaneously perceives* when perceiving the *shape-visible syntactic form* of the expression in a context. Just as the perception of color is infiltrated by background cues, so is what is said by an expression uttered. More particularly, whether what-is-said is contextually influenced depends on subtle facts about lexical meaning.[8] The main point is that what-is-said is *experienced* as a property of the

---

8   Azzouni discusses many examples. To give an idea of what he has in mind, consider the sentence 'that's a dangerous dog' (whose interpretation is radically different depending whether the sentence is uttered in a context where a barking big dog is approaching or in a context where the speaker and her interlocutor

expression, and as *unrelated to the speaker's intention*; contextual infiltration is never conscious. Conversely, what is implicated is related to the conscious detection of the use that a speaker makes of the expression. What we see is involuntary, affected by external cues that we may also perceive, but unrecognized – in the experience – as being affected by such cues. What is said, similarly, appears to us to be a monadic property of the uttered expression just as the color of an object appears to be a monadic property of it. Exactly as the mechanisms of perceptual psychology operate unconsciously, so the mechanisms by means of which a certain expression in a context is given a certain meaning for speakers/hearers are not accessible to introspection and are thus invisible to them. Speakers/hearers can see no conceptual connections between the meanings of public-language expressions and the intentions and behaviors of anyone any more than they can see conceptual connections between the actual properties of a rock (such as its hardness or color) and the intentions and behaviors of the person who has picked that rock up. We present here a relevant quote from Azzouni (2013), discussing the case on an ant involuntarily drawing intelligible signs in the sand (roughly equivalent to the sentence-tokens 'I am that I am,' 'I am happy,' 'he is happy'):

> These facts, about the involuntary imposition of meaning on shapes (I am that I am, I am happy, he is happy), are every bit as important *as their sibling facts in the visual domain*, as previously mentioned. In a sense, the perception of depth, in certain cases, and the perception of certain illusions […] are involuntary. Recall the view some sense-data theorists had, that the perception of depth wasn't 'given' but was rather a kind of rapid inference from the experience of flat sense data (which was given). This has a companion in the view that our perception of the meaning of shapes involves inferences from sheer embodied shapes to their meanings via background assumptions about language conventions or the beliefs and intentions of the producer of those shapes, and so on. Both views are utterly false to the phenomenology of the experiences in question (however true they might be—subdoxastically speaking). (Azzouni 2013: 80)

are trying to assess which dog breeds are capable of transmitting to humans some dangerous viral agent) and the sentence 'John went to the gym' (whose interpretation is radically different depending on whether the sentence is uttered in a context in which we are trying to assess whether John walked until a certain building or whether John was inside a certain building at a given time).

Whatever attitude the reader might adopt with respect to these ideas, we deem that the parallelism between Azzouni's attempt at reducing sentence-meaning to the unconscious operations of 'semantic perception' and Fodor's attempt at reducing concepts and reference to the operations of unconscious clusters of sensory features (within the 'perceptual circle,' at least as a first rough approximation) are in fact almost self-evident. What they have in common is especially the emphasis on the role played, both in perception and in cognition, by unconscious subpersonal mechanisms.

2.6 Only a modest role for content?
No, that's individualism!

Computational theories of vision (on the model of Marr's *Vision*; Marr 1982) raise the problem of perceptual content from a perspective that is particularly relevant for our purposes in this book. On the one hand, we have what F. Egan dubs the *Essential Distal Content View* (Egan 2010): the internal states, patterns and structures of the mind posited by computationalists are the domain of an interpretation function whose co-domain is objects and properties of the external world. This is by far the dominant view. On the other hand, we have Chomsky's *Methodological Eliminativism*, according to which the interpretation function can in principle be completely disregarded; what only matters is the causal properties of the syntactically characterized internal states of the mind (see also Stich 1983; Piccinini 2008). Chomsky's eliminativism holds in fact that the latter are the only legitimate targets of scientific inquiry, whereas the common-sense world of objects and properties delivered by our visual system typically respond to a fairly arbitrary variety of pragmatic factors and pragmatically relevant social interests. In this respect, Chomsky establishes an almost perfect parallelism between language and vision. In his view, the scientific inquiry into language as a cognitive system abstracts away from a naturalistic semantics, in the sense that the objects we talk about as referents of the linguistic expressions we use are epistemologically dubious, since they satisfy essentially pragmatic requirements.

Referential semantics, in the eliminativist view, does not exist as an independent field of scientific inquiry; to the extent semantics is allowed as a part of the scientific investigation of language, it is nothing more than a constitutive part of syntax (in the *broad* sense). In fact, what is *formal*

semantics if not a set-theoretic type of syntactic calculus in which the choice of the actual atoms of interpretation is entirely free? For the purposes of this calculus, nothing hinges on the choice of a particular set of objects and properties and Chomsky has repeatedly emphasized that lexical meaning (hence the objects and properties delivered by the relation of *reference*) is a mysterious galaxy that evades a scientific style of inquiry. We can of course use formal-semantic computations (among other things) to talk about the common-sense objects that populate the world of our experience; however, these objects are individuated by a completely independent (and scientifically irrelevant) net of practical interests and pragmatic considerations.

On these premises, Chomsky's interpretation of Marr's system is also radically *internalist*. According to the received wisdom, in Marr's system, the algorithms are intended to characterize the transition among the different syntactic states that define the visual computation, whereby causal relations crucially hold among these states. In early vision, for instance, channels and filters are activated that are sensitive to changes of light intensity. This set of computations is generally presented as delivering a representational primitive, that is, *edges*. But this should not be taken too seriously, since it corresponds – Chomsky contends – to an *informal* style of exposition.

It is useful to briefly consider the issue of edge individuation in some more detail. According to the canonical externalist approach to Marr's system, there is a *spatial coincidence assumption* to the effect that if a *zero-crossing* segment (a well-defined pattern of sudden light intensity change) is present in a set of channels of a contiguous range of sizes, and the segment has a very similar orientation in each channel, then the intensity change expressed by that set of zero-crossings is due to a single physical phenomenon, typically the presence of a real *physical edge* (Burge 1986). According to this view, then, certain computational states directly correlate with certain representational primitives (local contours, edges, shadows, oriented surfaces, etc.). It is in this sense that this view qualifies as externalist.

Crucially, however, this is not the way Chomsky thinks of Marr's theory of vision. As already mentioned, for Chomsky the received wisdom is simply the consequence of an *informal* style of presentation, which facilitates the discussion of the relevant algorithms. *A priori*, there is no reason, in fact, why a specific pattern of zero-crossings should be related to some particular

physical phenomenon in the environment. What really matters is the syntactic properties of the computational states and the mathematically defined transitions among the distinct computational states. Interpretation is thus a dispensable feature of Marr's computational theory of vision. If interpretation takes place, this is due to nothing else than our pragmatic disposition to put the system to use in some local context or other. Content is ultimately dispensable. Chomsky is a consistent eliminativist, both in language and in perception.

Whenever cast in the radical terms adopted above, Chomsky's eliminativism is intuitively difficult to accept (as well as its concomitant internalist interpretation of Marr). Most computationalists (for instance Pylyshyn 1986) would not only contend that representational states of the mind have a causal power in virtue of their syntactic properties, they would also crucially contend that there is a *naturalistically* defined relation between those mental structures and the external objects and properties into which they are mapped. In their view, interpretation is essential to the computational states of the mind, since there is a causal relation between the properties of these states and the (ultimately physical) properties of the distal objects that these states are meant to represent. Even more compellingly, perhaps, the computational states that define vision are inherently characterized as information-extraction devices (Matthen 2005), intended to provide a veridical representation of some distal properties. Granted, there is no direct correspondence between a set of objects visually experienced as having constant size and shape and the distribution of light on the retina. However, the point is exactly this: information-extraction (say, zero-crossing computation) is inherently *intentional*, for the very reason that it strives, since the very beginning, to the individuation of some representational primitives (edges in the case of zero-crossing computation, constant size and shape in other cases), reflecting some distal properties of the physical world.

So, is Chomsky's eliminativism easily dismissible? Is the role of content in computational theories of vision firmly established? Is Chomsky's strongly internalist interpretation of Marr nothing else than an unwarranted extension to vision of the view he holds for language, that is, the view according to which semantics resolves in a syntactically defined set-theoretic style of computation?

Things are not so straightforward. An interesting defense of Chomsky's eliminativist view is found in Egan (2010). At first sight, Egan's approach is less radical than Chomsky's in that she thinks that representational content is a required component of any adequate computational theory of vision, since it is required to define the theory's *explanandum* – the domain of cognitive phenomena that define its explanatory domain. Egan's position is that representational content is not part of the definition of a theory of vision, it is rather nothing more than a pragmatically motivated 'intentional gloss.' More precisely, representational content is a gloss that allows the computational theory to address ecologically motivated local explanatory interests. Cognitive interpretation (the interpretation function that strives towards the individuation of representational primitives, distal objects and distal properties) has thus no *individuative* value for the computational theory. Vision needs not be *interpreted* vision. All we should accept of the externalist standpoint according to which representational content is *individuative* of the theory, is that representational content is useful to define a domain of cognitive objects to which the theory *may* apply. In a way, the intentional gloss is necessary to show that a certain theory of vision does well its job in terms of the pragmatic interests of some ecologically defined local audience.

In this respect, we can also say that the objects vision revolves around are the common-sense objects of our visual experience. Epistemologically, however, these objects have no value, as held by Chomsky. There is a sense, thus, in which the only ontologically relevant interpretation function is not the function that maps computational states of the mind into cognitive objects and properties; rather, it is the function that maps these patterns and structures of the mind into an abstract domain of mathematical objects (Marr's algorithmic level). Since the latter have, by definition, no causal power, the possibility of a naturalistic semantics in the domain of perception is forcefully denied.

Clearly, in spite of her moderate tone and her apparently balanced standpoint, Egan's view is very close to Chomsky's position, of which it emphasizes in fact the potentially 'idealistic' contours: the reality that vision discloses is not the domain of the ontologically dubious cognitive objects and properties we experience in vision and talk about in language; it rather coincides with the ontologically established domain of abstract

mathematical objects. The following quote is in this respect straightforward enough:

> A second (related) virtue of taking the mathematical interpretation, rather than the cognitive interpretation which specifies representational content, to be decisive for computational individuation is that a mathematical characterization of the mechanism advances our understanding of the mechanism by subsuming it under a description that we independently understand. We have an independent understanding of such mathematical processes as vector subtraction, Fourier transforms, and so on. This view emphasizes the fact that computational theorists have a particular set of tools—mathematical functions—and that the distinctive contribution of a computational account is the successful application of these tools to a cognitive domain. (Egan 2010: 258–59).

We should notice that the most revealing part of this quote is the final line, where the computational theory of vision is taken to successfully apply to a *cognitive* domain of objects. In other words, these objects no longer constitute the *external* correlates of internal states of the mind, as they emerge based on the success of the mathematical computations in delivering, through the necessary stage of representational primitives, distal objects and distal properties. These *cognitive* objects are just pragmatically motivated objects, largely overlapping with our common-sense understanding of the world, though they are undoubtedly useful in putting to the foreground the success of the theory in its contextual, local applications.

In order to increase the plausibility of the 'modest role' she would assign to representational content in a theory of vision, Egan makes effectively use of her 'Twin-Visua' thought-experiment. Here is how we might summarize it. Let's dub our visual system (as it developed on Earth) *Visua*, and let us assume, for the sake of the argument, that there is a distal property C_1 that it successfully tracks – say, changes in depth. On a different world, the very same computational mechanism appears – let's dub it *Twin-Visua* – defined by the same set of computational structures and states and by the same mathematically expressed transitions among those structures and states. The two mechanisms have thus the same discriminative and recognitional capacities. The difference is that Twin-Visua did not develop as an adaptation on the other world, that is, there is no distal property C_1 to be tracked by Twin-Visua on the other world. Clearly, nothing excludes that Twin-Visua, though physically and computationally indistinguishable from Visua, accidentally tracks, in the second world, a distal property C_2

which is crucially distinct (in fact, possibly quite different) from the distal property C_1 tracked by Visua on Earth. Actually, it is quite possible that the set C_2 of distal properties accidentally tracked by Twin-Visua in the alternative world be so different from C_1 that Twin-Visua does no longer qualify as a visual system (but, say, as a tactile system) in the alternative world. This would obviously depend on the fact that the distal/physical conditions and the laws of nature in the alternative world are different from the conditions on Earth. Under these assumptions, there are two distinct interpretations of the very same computational mechanism (C_1 and C_2), responding to the different environmental conditions in which the computational mechanism is made operative in the two worlds. Clearly, these two interpretations are irreducible. It would be plainly wrong to assume that the computational mechanism is interpreted in a unique (presumably more abstract) manner that subsumes the two original interpretations. To shortly see this, simply suppose that this more abstract interpretation tracks the distal object C_3 (by definition, neither a visually nor a haptically defined object). It is evident that Visua cannot be assumed to track C_3 on Earth, since this would not adequately capture the essentially visual nature of the object tracked by Visua under typical conditions on Earth, for instance the fact that it is a visually experienced object. It follows that there is not a unique interpretation for the computational states and structures of Visua, which demonstrates in turn – according to Egan – that (i) representational content is contextually defined, based on pragmatic considerations, and (ii) representational content cannot be an essential part of computational theories of perception.

So, these being the essential ingredients of the story, what do we make of Egan's thought-experiment of Twin-Visua? We believe that Egan's argument does not go through. The crucial point of criticism is not the independence of the algorithms from representational content. This is a firmly established thesis. However, this thesis has rather trivial consequences – and certainly not the consequences that Egan would like to draw from it. To be sure, we can assign different interpretations to a syntactic system, as well as we can use a certain set of mathematical devices to explain different sets of phenomena at the physical level. More particularly, logical systems can be interpreted according to non-canonical models. This does not prevent these systems, of course, from being developed in order to correctly and

adequately capture a given set of canonical interpretations. In other words, this does not prevent these systems from being strictly tied to the set of interpretations they canonically express. In the case under scrutiny, we cannot certainly exclude, as we have seen, that Twin-Visua arose as an effect of a sudden mutation, to the effect that the way it reliably tracks some physical properties in the alternative world is just a fortunate coincidence. In fact, it is equally plausible that Twin-Visua, as a non-adaptive computational mechanism, does not track anything interesting in the alternative world.

The critical point, then, is simply that the way in which Visua or Twin-Visua relate to the distal properties they actually track is an empirical property of these computational systems, as such worth investigating. In our world, for instance, the visual system plausibly developed to detect and represent physically significant correlates of the computational states that define its operations. To assume that interpreting the algorithms is only a matter of pragmatic interests introduces an anti-naturalistic, illegitimate divorce between the mind and the external world. To cut a long story short, the mind's abstract computations represent, within the visual system, a solution to the problem of obtaining a veridical representation of the world. We believe that this is the null hypothesis, and that departing from this assumption should be justified. According to Burge, 'individualism' in psychology is taken to be the view according to which there is no deep *individuative* relation between an organism's mental states and its environment. To assume that interpreting the algorithms is only a matter of pragmatic interests is thus replicating the classical mistake of individualism: it precludes a naturalistic style of explanation in semantics, according to which the computational states of the mind originate as a solution to the problems posed by the specific challenges that an organism's mind has to face in a specific ecological setting. In vision, representations with intentional content like generalized cones – the prototypical form of three-dimensional objects – and edges, oriented surfaces (on a par with, more generally, any kind of representational primitives) reflect the presence of an *individuative* relation between mental patterns and computations, on the one hand, and physical properties of the external environment, on the other.

In a nutshell, Egan's point is that even if the organism's physical mechanisms (say, Visua and Twin-Visua) remain the same (hence they qualify as a unique mental state), the kind of intentional representations

that vision theory would attribute to them would be different if the environmental conditions were different, even slightly different. This point is granted, but only under the compulsory interpretation that if the environmental conditions are different, so are the constraints or, better to say, the formation laws that govern the relation between these environmental conditions and the organism's mental states. What follows is thus, in T. Burge's words, that 'if the regular, law-like relations between perception and the environment were different, the visual system would be solving different information-processing problems' (Burge 1986: 34).

If more interpretations are possible, it does not follow that a computational theory of vision is inherently non-intentional. What is lost in Egan's notion of 'intentional gloss' is the notion that mental states are not freely available computational patterns that are only in need of some contextualization. They do not simply deliver a pragmatically biased interpretation of the specific context in which they are put to use. This is nothing else than the classical error of Burge's 'individualism,' and represents – we believe – a strong departure from the scientific style of inquiry in evolutionary biology, even with respect to the more recent stream of thinking in evolutionary developmental biology.

As Burge emphasizes, it is investigation of general physical facts that allows interpreting the computations performed by the visual system as a solution to the problem of extracting reliable information about the physical properties that are ecologically relevant for the organism. This explains the relative success of visual representations: they are veridical under standard conditions, that is, they are likely to correlate to a specific set of ecologically relevant distal conditions. The conclusion is thus that visual systems are inherently representational – not in the sense that they cannot be conceived and described independently of the interpretation they standardly receive, but in the sense that these interpretations are just not pragmatically useful additions to the system; rather, they are ecologically responsible veridical representations of the physical objects and physical properties in the environment. This seems, after all, what Marr had in mind:

> The method is again to identify general physical conditions that give rise to a visual process, then to use those conditions to motivate constraints on the form of the process that, when satisfied, will allow the process to be interpreted as providing reliable representations of the physical environment (Marr 1982: 116).

The point at which we diverge from Egan and strongly concur with Burge is thus essentially the following: *cognitive interpretation functions* within a theory of vision (conceived as a system of causally interacting computational structures) have an essential, ineliminable, epistemological value. They do not simply provide ecologically opportunistic routes to action; rather, they deliver ecologically useful representations of the world. Perception is inherently representational, hence inherently intentional. Egan's 'modesty' is not a satisfactory way of representing representational content in perception.

There is a last interesting consideration to draw from the whole set of observations above. Chomsky's strongly internalist positions are often motivated, with respect to language, by the concomitant position that formal semantics, as seen above, is nothing else than an abstract calculus based on set-theory. As is the case for Egan, there is then no need for a naturalistic semantics: which kind of objects this set-theoretic calculus quantifies over can be left entirely indeterminate. In fact, as also already seen, lexical and conceptual atoms raise, in Chomsky's view, a whole set of scientifically intractable problems. This approach obliterates the strong naturalistic foundations of Fodor's view (Fodor and Pylyshyn 2015): according to Fodor, percepts/concepts arise in the mind based on causal relations with the world, originally, and crucially, within the perceptual cycle. From Fodor's perspective (as well as from Burge's), vision is not a self-referential computational device, it revolves around the individuation of representational primitives and is endowed with an intrinsic intentionality, tracking distal properties and providing (under typical circumstances) veridical representations. Percepts and concepts are thus not simply a set of virtually intractable problems in what concerns their relation to the external world, they constitute in fact an arguably viable solution to the problem of veridically representing the world, that is, the problem of providing correct (though not always accurate) representations of the external world (ultimately, the physical world). In a sense, then, it is Chomsky's radically internalist view of perception, whereby no role is set for representational content as one of its essential ingredients, that prompts Chomsky's internalism about language, and, more particularly, Chomsky's skepticism about the notion of reference in semantics.

On these grounds, the conclusions we have reached above impact language not less than perception. *Pace* Chomsky, abstract computations,

both in vision and in language, need not be divorced from reality. In perception, they arguably provide subpersonal algorithmic routes for the individuation of distal objects and distal properties; in language, they provide subpersonal algorithmic routes to make a sophisticated cognitive use of the distal objects and distal properties individuated by perception. Here lies – we believe- the scientific core of a naturalistic theory of meaning.

2.7 Some conclusions on perceptual content

Perception is no less essential to our mental life than so-called higher-cognition (language, thought, reason) is. The perhaps most pervasive account of perception as the original source of our cognitive life and as primitively and irreducibly linked to perceptual content is Burge's *anti-individualism* (see the discussion in Chapter 1), a theoretical stand rooted in evolutionary and phylogenetic considerations and fed by the experimental results obtained in perceptual psychology and, more generally, cognitive neuroscience. We firmly believe that Burge's ideas deserve a prominent position within the present state-of-art in the philosophy of perception.

According to anti-individualism, if there are perceptual states, their function is to be *veridical*. This clearly entails that the representational content of perception must be based on the presence of *conditions on successful representation*. In other words, perception is inherently rooted in veridicality, in the sense that there must be objects causing the relevant perceptual states and these objects must be endowed with the properties (attributes) that are ascribed to them.

At the same time, Burge's view should be sharply distinguished from Armstrong's *causal account* of perceptual knowledge (Armstrong 1973), which abstracts away from perceptual experience and establishes a direct causal link between unconscious perceptual knowledge (as in Dretske's simple seeing or in blindsight) and epistemic knowledge. Armstrong's solution simply puts aside the issue of *justification*. For Burge, as in all externalist/reliabilist approaches, the issue of justification is the core issue and can be formulated as follows: the consideration of the means by which perception becomes objectively reliable, and becomes in fact the foundation of objective knowledge, and of how reliable perceptual beliefs are supported by reliable perceptual processes. From Burge's perspective, the basic representational function of perception is accurate representation of the subject

matter that is sensed. More particularly, a certain perceptual state can be effectively described as the ascription of a property F to an object *o* (a representational state is necessarily a propositional state). Perception is thus the first non-deflated kind of representation and there are norms for correct representation of perception, as there are norms for the correct representation of belief and belief states. The formation laws – that is, the laws of correct representation in perceptual psychology – that are operative in perception can be conceived as algorithmic operations performing the mapping from sensory registration to veridical perception states. These formation laws include, among others, principles such as *convergence*, whereby distance and location are calculated on the basis of vengeance and version angles, by using the geometry of binocular vision, and *constancy of lightness*, based on the separation of reflectance and illumination.

The kind of representational objectification supported by the systems of perception and belief, which gives rise to the sort of intentional action we are used to, is clearly lacking in the kind of low-level action engendered by registration of information based on response to mere bodily stimulation. For instance, the pattern of stimulation stemming, in vision, from the original registration of the input, in terms of spatial patterns and spectral properties of the light, has a mere information-theoretic relevance, but no perceptual import. In fact, one of the core notions in the perceptual psychology of vision is that *proximal stimulation* underdetermines perceptual states, since there are many possible causal antecedents of proximal stimulation that are compatible with a given proximal state. This entails that the visual system must be equipped with the capacity on making effective bets while selecting the *distal stimulus* that is most likely to have caused the relevant proximal stimulation. In this way, the many-one mapping from distal stimulus to proximal stimulus, as well as the concomitant one-many mapping from proximal stimulus to distal stimulus, are regularly neutralized.

Here, Burge's anti-individualism manifests, as hinted at above, its strong evolutionary roots. The formation laws of perceptual psychology are the result of the adaptation of living organisms to the features of the environmental *niche* in which they thrived. These laws produce an output – a relevant perception state that is represented in the system – but the laws themselves are not represented – an important point we already made above

concerning the fact that modalities of information-extraction in perception are both *conceptual* and *unconscious* to the perceiver.

It is this feature that most profoundly defines Burge's anti-individualism: the cognitive principles that build a perceptual representation are not part of that representation. Do these principles simply depend from biological function? The answer is negative. As such, perception is independent of biological function: Burge emphasizes, contra Millikan, that even non-veridical representations, as in animal-danger perception, can still be biologically functional. Veridicality and, more generally, the representational power of perception, cannot be fully explained in evolutionary terms. In this respect, notice that evolutionarily-induced responses to patterns of sensory registration are not necessarily endowed with a representational content. Take the relatively often discussed case of bacteria reacting to magnetic fields linked to the presence of beneficial de-oxygenated areas. What is going on here is a pattern of sensory registration attached to an automatically induced bodily response. The question about perceptual content is conceptually independent. Perceptual content is absent if there is no representation beyond sensory registration; here, action is explained away in terms of sensory discrimination (sensitivity to magnetic fields), not in terms of the capacity of representing a locus characterized by a property (oxygen-poverty) that is beneficial to the perceiving organism (in other words, no object o is represented as endowed with attribute F). Perception fully expresses predication, and predication, as the representational kernel of the higher cognitive functions typically encoded by language, predates language and is not a consequence of language.

Chapter 3   The labyrinths of reference: Mirror effects in language and vision

3.1 Introduction

One of the major challenges in the scientific study of language is provided by the issues surrounding the notion of reference. To present and discuss this notion and the challenges it raises in a way that is functional to the goals of the book, we will start by considering a specific theoretical approach. This approach corresponds to a well-established tradition in the philosophy of mind – largely subsumed by formal approaches to the semantics of natural language rooted in Gottlob Frege's work – and maintains that when we use language to ascribe an attribute to an object, we produce linguistic expressions in which a *concept* is applied to an object. It further assumes that objects are also identified through concepts. More particularly, objects are always referred to by means of descriptions of the form 'the *F*,' that is, the concept of being uniquely *F*. Even proper names, despite their apparent referential properties, are in fact understood as concealed descriptions. So, when a cognitive agent ascribes the concept 'red' to object *o*, it does so by conceptualizing the object as well (as, say, 'the flower I see now'). This view, of course, corresponds to an 'internalist' understanding of the notion of meaning. Propositions are made of concepts, not of references. They are mental representations consisting in a sort of conceptual composition.

As a point of historical clarification and to avoid misunderstandings, it is important to keep in mind that this internalist view should not be identified with the work of Frege or Russell, who explicitly adhered to a rigorous form of Platonic realism concerning notions such as 'concepts' and 'ideas.' It is also true, however, that both Frege and Russell crucially contributed to the detection and promotion of some of its crucial ingredients, as attested by the fact that some of the main early advocates of semantic externalism, such as Kaplan and Putnam, explicitly identified the form of internalism we are now discussing as 'the Fregean view.'

Be as it may, the form of internalism that is the object of discussion of this chapter is grounded on the insight that propositions are in the mind. More technically, propositions are intended to model cognitive operations that are defined in abstract set-theoretic terms.

Propositions contain functions from worlds to objects – such as definite descriptions of the form 'the *F*' -, functions from worlds to set of objects – such as the predicate 'red' -, and they are themselves functions from worlds to truth-values. The notion of proposition that emerges models a cognitive operation that is defined in abstract set-theoretic terms. Hence, propositions are not entities external to the mind, they do not coincide with the worldly states of affairs in which, say, the unique object *o* that has property *F* also satisfies the property of being red. They are rather made up of concepts, not references, and concepts express cognitive operations of the mind.

We should also observe that, within this framework, meaning is a higher-order property. It is not only inherent to the mind, it also crucially expresses higher-order operations in the mind, that is, a complex collection of set-theoretic operations by which object-level properties are applied to objects – as when we say 'This is a flower' – property-level properties are applied to object-level properties – as when we say 'every rose is a flower' – two object-level properties combine together – as when we say 'red flower' – and so on. In this sense, this version of semantic internalism somehow reflects, in semantic terms, Burge's notion of individual representationalism, as discussed in Chapter 1. The reason is that all this set-theoretic apparatus is supposed to come along in one fell swoop. In fact, as we have seen in Chapter One, the very notion of having concepts has been long held as the trigger of a sort of 'snowball effect': a cognitive agent is not deemed to be able to entertain the concept corresponding to 'red' if she does not also entertain the concept of 'concept' and if she is not able to use the concept 'red' in a large variety of non-trivial functional applications, such as 'whatever I see is red' or 'I saw something red.' Propositions are made of concepts, and the possession of concepts arguably entails the possession of an intricate, in principle unconstrained, set-theoretic machinery.

Thinking – as well as speaking – must then involve a cognitive apparatus that goes far beyond perception, especially if perception is interpreted as the passive reception of sensory-features. Maybe honey bees see the world in a way somehow phenomenally comparable to the way *we*, humans, see the

world, but it is *only us* who can use visual perception as the basis of conceptual representations that are rich in content and contribute the building blocks of a complex array of cognitive computations. The reason is, as we saw, that it's only us who have concepts. Perception is simply the passive database to which conscious operations of the mind apply. And even if there is perceptual content, and this content is essentially representational, there is no way perception can pretend to be something more than a sort of proto-cognition. Whenever perception appears to us more complex than it should be, a widespread tendency among philosophers and linguists is to discharge this complexity on issues of 'cognitive penetration.' Once more, it is ideas (reason, cognition) that makes the world intelligible, not the subjective mass of analogical sense-data we are exposed to.

Under similar assumptions, it is not difficult to incur in the fallacy of identifying the mental capacity for concepts and propositions with the capacity for grammar (as W. Hinzen actually proposes; see the discussion in Chapter 1): if concepts and propositions are higher-order cognitive skills that set us apart not only from bees but also, and most notably, from the other primates, what better proof of that than the identification of fully-fledged thinking with grammar, especially since grammar is – rather uncontroversially – a distinctive feature of humans?

This fairly narcissistic view of the capacity for meaning and reference as embedded in the human capacity for language is fundamentally wrong – for a number of reasons that we are going to review in some detail.

3.2 Reference in language and perception

To begin with, the view of perception content that we have endorsed in Chapter 1 and further discussed in Chapter 2 makes concepts quite less detached from perception than commonly believed by linguists, a position, as we saw, strongly endorsed by Fodor (see the discussion in Chapter 2). We have also seen that perception is endowed with combinatorial properties. This does not mean, of course, that perception is endowed with the full range of combinatorial possibilities that are made available by syntax in language; however, it minimally entails that the divide between language and perception cannot be identified with propositional thought. Propositional thought is already a property of perception. In fact, it is a property of perception no

less than it is of language. This is tantamount to asserting that one of the main tenets associated with the internalist view of propositions discussed in the first section above is not warranted: even if propositions are made of concepts, this does not mean *per se* that propositions (and, with them, the capacity for veridical representation) are a prerogative of language as a higher-level cognitive system.

To examine this point in some more detail, let us go back to Matthen's view, as presented in Chapter 1. Sensory-features are not only classificatory, they are also inherently propositional. They are wired-in classificatory devices whereby the response conditions to sensory inputs are defined propositionally ('this is red,' 'this is a flower'), assigning to these sensory inputs discrete (Boolean) or, more realistically, probabilistic (Bayesian) values. The classificatory power of sensory features is based on a set of formation rules, performing the mapping from analogical databases such as retinotopic images to some final perceptual states, whose order of computational complexity is not lower than that proper to the view of concepts as mathematical functions. In fact, sensory-features corresponds to computationally complex, evolutionarily induced patterns of information extraction from analogical databases. As such, they incarnate the notion of *perceptual concept*. Though the full array of concepts made available by language can be the result of the manipulation of the original perceptual content (Matthen 2005 offers in this regard interesting exemplifications with color terms in language), this observation is perfectly compatible with the claim that *linguistic* concepts are strongly grounded on *perceptual* concepts.

Still, one might object that what matters is not the fact that concepts correspond to computational devices, but the way in which concepts are assembled together. In this respect, an important merit of the internalist view, in fact one which is genuinely rooted in Frege's legacy, is its emphasis on *compositionality*: the semantic value of propositions as functions from worlds to truth-values is compositionally derived from the semantic value of the conceptual components of the proposition. For example, in the case of sentence *S* 'the flower I see is red,' the semantic value of *S* is compositionally derived from the semantic value of 'the flower I see' and 'red.' This logical style of composition is crucially based on higher-types (in fact, potentially unconstrained higher-types).

The question becomes then the following: Does the view that propositions are rooted in perception compel us to reject the view that propositions, as expressed by language, are set-theoretic constructs built by recourse to higher types? The answer – we believe – is that there are no compelling arguments for rejection. After all, a set-theoretic framework for assembling concepts guarantees compositionality and allows recursion. It explains why quantification (predicating properties of properties) is such a natural device in syntax; why propositional representations are possible not only in the absence of a direct perceptual source (*weak decoupling*) but also without being directed towards action (*strong decoupling*);[9] why the basic ingredients of thought, that is, propositions themselves, can easily be made the object of functional application in syntax (or the language of thought), as when we deal with propositions, in language, as the semantic value of the clausal complements of verbs of propositional attitudes and as the values of certain quantifiers, and so on. It is in fact quite plausible that having syntax is strictly related to having this capacity of linking concepts to abstract set-theoretic values.

Consider that, in this respect, a defining feature of human language syntax is the distinction between *lexical* and *functional* categories. According to some influential syntactic models, as in the cartographic approaches inspired by Cinque's templatic model of functional hierarchies (Cinque 1999), the universally defined array of functional categories that constitute the 'extended projection' of lexical categories such nouns and verbs is the backbone of syntax in its 'narrow sense.' But what are these functional categories exactly? According to Fintel and Matthewson there are two basic types of expressions:

> Expressions that refer to entities or situations and expressions that denote predicates of entities. Functional categories have high semantic types, beyond these levels of entities, situations, and predicates (Fintel and Matthewson 2008: 159).

Let us suppose, following Fintel and Matthewson's suggestion, that functional categories correspond to logical terms whose interpretation is

---

9   We are using the notions of weak and strong decoupling as they are defined in Reboul (2017).

fully – and universally – determined by grammar, whereas lexical terms are the non-logical words, whose meaning is underdetermined by grammar. This leads to the restrictive hypothesis that functional terms are those terms that express high-type meanings. If these high-type meanings are assembled by grammar according to rigid templatic structures, as in cartographic approaches to syntax, it is these high-type operators that constitute the semantic glue that keeps lexical meanings together and effectively lead to the calculation of complex compositional interpretations.

In a sense, this is a way of technically implementing Saussure's basic insight that language hides a rich texture of grammatical relations and that it is these grammatical relations that are responsible for the way meaning is compositionally encoded. The structuralist tradition capitalized on this insight to derive the notion of meaning, in its entirety, as an emerging property of grammatical relations. We know now, however, that language is not really like music, where what characterizes a melody is not the exact notes that compose it, but, rather, the relations of time and pitch that hold between them. In language, grammatical relations are not all there is to meaning, as demonstrated by the fact that knowing the properties of a complex grammatical structure is never a sufficient condition for inferring its meaning. We can appreciate this point by considering L. Carroll's Jabberwocky or sample sentences as the notorious 'The gostak distims the doshes' (Ogden and Richards 1923), which seems partly intelligible to most English speakers though, of course, not entirely so. Obviously, what is also required for a full understanding of linguistic expressions, no matter how rich and complex in functional structure, is the meaning of the lexical terms that occur in them. When it comes to lexical terms,

> […] learning their meaning is determined causally, and will be affected by experience, perception, knowledge, common-sense, etc. But none of these factors is relevant to the meaning of quantifiers. The child has to learn the content of the lexical entries for the non-logical terms, but this is not necessary for the entries for the logical terms, for they are given innately (May 1991: 353).

As we have seen in Chapter 2, it is these lexical terms that correspond, according to Fodor, to the atomic concepts that are causally determined, at least to some extent, within the 'perceptual circle.' However, from May's quote above we infer that for the standard linguist it is not only perception, but also the rest of cognition (common-sense, knowledge, etc.), that

is responsible for the way concepts are fixed during the learning process. This is of course not surprising at all, since we have emphasized, in the Interlude, that one of the open problems for the Fodorian view is that many concepts are not realistically fixed by the unconscious action, within the perceptual circle, of certain clusters of sensory features, but correspond, rather, to abstract entities that are not immediately available within perception. It is this problem that probably makes the so-called *theory-theory* of concepts potentially attractive (Carey 2009). According to this view, concepts store all kind of knowledge that can be relevant for the members of the concept's extension. However, reducing concepts to categorization (as in Millikans's 2000 approach) or attributing to them the complex functional and explanatory structure of theories also leads, as discussed in Chapter Two, to a number of unpleasant consequences, especially since recognizing a member of the concept's extension as belonging to the relevant category often fairly exceeds the cognitive potential of the conceptual mind. In other words, to have a concept is not the same as being able to establish which entities fall in the extension of this concept or to be aware of the full functional and explanatory power associated with the concept.

What this suggests is that Fodor's approach is essentially correct: concepts are causally determined by the action of clusters of sensory features within the perceptual systems. This is an unconscious process that involves neither categorization nor conceptualization: for the perceptual agent the concept is, in principle, simply a computational atom that enters as such the systems of language and thought. According to Fodor, to possess a concept is to be able to use it as an atom within a digital computation.

It is here that the role played by functional structure and functional categories in language may be relevant for the empirical appeal of this view. After all, possessing an abstract concept (say, 'electron') is less based on the encyclopedic information or structured knowledge that we attach to it than on the appreciation of the formal relations that it comes to entertain with other concepts by means of its use in language. When we say, for instance, that 'Every electron has an electric charge,' it is language itself, through the texture of logical relations induced by the functional apparatus, that substantially contributes to fleshing out the meaning that we associate to the word 'electron.'

From this point of view, it is also appealing to make use of J. Azzouni's notion that linguistic meaning is directly and unconsciously 'perceived,' in the sense that the interpretation we associate to a certain sentence token, as uttered in a given context, consists in the unconscious activation of our perceptual and cognitive systems, whose operations, streamlined by endogenous attention, automatically and mechanically exploit all the available linguistic content and contextual clues associated with the utterance (Azzouni 2013). Linguistic 'perception,' in the sense of unconscious processing of the logical relations induced by functional structure in language, is thus an important part of the story and may play a causal role in explaining why the human conceptual apparatus quickly extended beyond the 'perceptual circle' in the narrow sense.

In a nutshell, what we are proposing is that language itself (or, better to say, functional structure as the logical backbone of language) becomes, for humans, part of something we could define as an 'extended' perceptual circle, which contains a fixed system of logical relations based on high-types. In this way, both perception and language are based on the presence of a biologically and phylogenetically determined underlying rich structure, which corresponds, in the case of language, to the deployment of a rich texture of set-theoretic relations, which are encoded, to a significant extent, by functional structure.

As rooted in perception, propositions are classes of cognitive events whereby, typically, a property is ascribed to an object. They are representational cognitive events associated to veridicality conditions, which in turn depend on the principles regulating correct perception as established by the empirical research in perceptual psychology. However, once transferred into the language system, propositions are mapped into set-theoretic objects, and encode as such a rich set of formal operations which is independent – as far as this computation is concerned – from the original perceptual conditions of content representation.

This is why – we believe – Chomsky's claim that formal semantics is part of the – broadly conceived – syntactic computation is essentially correct. After all, it is syntax that incorporates lexical words – conjuring up a rich texture of perceptual and cognitive relations (Pietroski 2018) – into the network of set-theoretic relations encoded by functional structure.

One important point discussed in Chapter 1 was that the set-theoretic notion of proposition cannot be used to defend the idea that this set-theoretic definition has a foundational value. In fact, a set-theoretic definition of proposition leaves it unsettled *why* a proposition has the veridicality conditions it happens to have. More precisely, as Scott Soames has convincingly argued, any time we try to *derive* the role of propositions as truth-bearers, we unavoidably presuppose the concept of proposition. Remember the core facts. If the basic propositional format consists in ascribing an attribute to a particular, the veridicality conditions associated with this ascription must involve the whole set of relations between environmental conditions and proximal objects and between proximal objects and the final perceptual state. If I ascribe, in vision, the attribute 'red' to a certain flower, whether my ascription is veridical depends on whether the laws of optics and the laws of formation of perceptual psychology have applied correctly. If something goes wrong with the laws of optics, I might be victim of an optical illusion; if something goes wrong with the laws of formation, I might be victim of a hallucination. In either case, my propositional ascription of the attribute 'red' to the relevant flower is not veridical. This is what justified Soames' claim that propositions cannot be foundationally defined as abstract logical structures but should be defined as types of cognitive events of property-ascription. If we omit to do so, we will not be in the condition to explain how we can *correctly* represent the world in perception and cognition.

What we have just seen is that it would be equally wrong to infer from these observations that propositions cannot behave as set-theoretic objects. To the contrary, dealing with propositions as set-theoretic objects establishes the bridge to a quite successful style of cognitive computation: the essentially syntactic computation that characterizes the systems of interpretation in language.

Our final claim is thus the following: the fact that propositions are rooted in cognitive acts performed by a cognitive agent does not prevent propositions from being endowed with further higher-order properties, aimed to make them computationally tractable within the systems of language and, arguably, thought. In this vein, it would be preposterous to deny that when propositions are realized linguistically as sentences, they

inherit properties *independent* of perception, like hierarchy, recursion, locality and, crucially, a set-theoretic style of composition. In fact, as hinted at above, it is quite possible that the set-theoretic combinatorics and the properties of syntax are intimately connected as a foundational property of language, in accordance with Richard Montague's basic insight.

Nonetheless, the foundational insight is that the set-theoretic notion of proposition is not the primitive one and that what ultimately ensures the possibility for sentences in language to be veridical is the notion of proposition that already emerges in perception, rooted in the veridicality conditions associated to individual cognitive events of property-attribution.

A further caveat is required. It would be fallacious to grant set-theoretic computations the property that the notions of truth and meaning are definable, and become accessible for a cognitive agent, only thanks to them. Truth and meaning, regardless of whether they are actually recognized as such by the agent, already emerge from cognitive acts of perception, and are thus logically and empirically prior to language. This point was already discussed in Chapter 1. To make the reader fully aware of this important implication, here is an effective quote from Soames:

> Unlike the Platonic epistemology required by traditional theories of propositions, the present account demystifies our acquaintance with, and knowledge of, propositions by taking both to be grounded in concrete cognitive experience. The explanation starts with the idea that we predicate properties of objects in cognition and perception, thereby entertaining propositions. This is done before we have the concept proposition. Focusing on similarities and differences in our experience, we eventually acquire the concept, making propositions objects of thought and subjects of predication. This allows us to acquire the notion of truth, in part by being given numerous examples – 'the proposition that o is red is true if o is red, the proposition that o is red isn't true if o isn't red,' etc. – and in part by coming to recognize the general point that a proposition is true iff things are as it represents them to be. Given truth, properties can be conceptualized as things true of other things. (Soames 2015: 104)

On these grounds, there is no point in the claim that an internalist view of concepts/propositions supports the 'narcissistic view' of the cognitive capacity for reference, holding that this capacity is a function of the higher-order cognitive systems of language and thought. What we should conclude is, rather, that the cognitive capacity for reference is deeply rooted in the representational systems of perception, though this does not exclude that concepts and propositions be mapped into set-theoretic objects as the result

of the *syntactic* style of computation which is proper to the independent – and extraordinarily effective – cognitive module constituted by language.

3.3 The conundrums around first-person reference

In the preceding sections, we have established that propositions are referential because they are representational, not because they are the result of a set-theoretic mode of concept composition based on logical-type recursion. That is not the whole story. There is yet another reason why the narcissistic view is untenable. There is strong evidence that propositions are not always made up of concepts when propositions are expressed in language. It seems that some of the linguistic expressions that make up propositions do not express concepts at all. Rather, they are *directly referential*, that is, they refer to particulars without the use of a descriptive apparatus.

First of all, as is well-known, Kripke has provided compelling arguments to the effect that proper names are not concealed descriptions (Kripke 1980). We will not review these arguments here, since they are fairly familiar and relatively uncontroversial. An important point that should be emphasized is that claiming that a name is directly referential is not the same as claiming that a name is *referentially rigid*. Rigid reference means that a linguistic expression picks up the same referent in all worlds, that is, what referent the linguistic expression refers to does not depend on which maximal state of the universe we assume we are in. In this sense, rigid reference is not a prerogative of names in language. A definite description like *'the successor of 0'* refers to the same number in all worlds (granting that we cannot imagine a state of the universe in which the successor of 0 is not 1). This description is a *de facto* rigid designator, since the fact that it refers to the number 1 is a metaphysical necessity, so to speak. Names are rigid designators *de jure*, in the sense that the fact that they pick up the same referent in all worlds is a property of language as a cognitive system (that is, a semantic fact) and does not depend on the metaphysical constitution of the world external to the mind. From this perspective, what should also be emphasized is that there are linguistic expressions – typically some descriptions – that are rigidly referential without being directly referential.

What 'directly referential' means is that names pick up their referent without the mediation of concepts: the name directly stands for the referent, so to speak. The most straightforward way of capturing this property

of names rests on the assumption that propositions containing names do not contain concepts among their constituents. Remember that the view of propositions as structured entities held it that propositions are a sort of logical duplicate of sentences. For instance, where we find a name in a sentence, we find a 'concept' in the corresponding proposition. Now, if names are directly referential, a sentence containing a name cannot correspond to a proposition in which the name provides the semantic value of a concept; rather, the name must be replaced, in the proposition, by the referent itself.

This consequence is effectively presented in this fragment from Kaplan's 'Demonstratives':

> Don't think of propositions as sets of possible worlds, but rather as structured entities looking something like the sentences which express them. For each occurrence of a singular term in a sentence there will be a corresponding constituent in the proposition expressed. […] The constituent of the proposition determines, for each circumstance of evaluation, the object relevant to evaluating the proposition in that circumstance. In general, the constituent of the proposition will be some sort of complex, constructed from various attributes by logical composition. But in the case of a singular term, which is directly referential, the constituent of the proposition is just the object itself. Thus it is that it does not just *turn out* that the constituent determines the same object in every circumstance, the constituent (corresponding to a rigid designator) just *is* the object. *There is no determining to do at all.* (Kaplan 1989: 185)

If this is the unavoidable consequence of direct reference within the framework of structured propositions, it is not of little significance. It entails that, even independently of the fact that propositions should be defined as cognitive events of property attribution rather than as abstract logical entities – say, structured propositions in the Russellian sense – we should not think of a structured proposition as necessarily made of concepts and only of concepts. Rather, there are circumstances, it seems, in which particulars external to the mind are direct constituents of a proposition. It is no accident that the view of names as directly referential has been used as an argument for *semantic externalism*, by Putnam and others (meaning is, after all, not 'in the head') and, more recently, as one of the arguments for the notion of *embodied cognition*, according to which the operations of the mind cannot be successfully described if one conceives of the mind as separated from the rest of the body and the external world.

It also turns out that names are, after all, just the starter. One of the best arguments for direct reference is in fact provided by

indexicals – context-sensitive expressions such as the first-person pronoun. Kaplan 1989 convincingly argued that some context-sensitive elements of language are directly referential. Suppose that D. Kaplan utters the sentence S = 'I am the author of "Demonstratives"' in a context c (that is, at a given time and while being in a certain place). What is the meaning of S? That is, which proposition does S express? Intuitively, one might be inclined to say that it is the proposition 'the utterer of S in c is the author of 'Demonstratives,' whereby the first-person pronoun 'I' is replaced by a context-sensitive concept, that is, by the description 'the utterer of S in c.'

This type of solution is clearly in the spirit of the semantic framework outlined at the very beginning of this chapter and, at first sight, might seem the best approach to tackle the matter. The semantics of 'I' is provided by the context-sensitive description 'the speaker in c;' the referent of the first-person pronoun in a sentence S is whoever utters S in a given context. So, if S is uttered by someone else than D. Kaplan (and at some other time and place), the referent of 'I' in S would not be Kaplan, but the speaker in this newly defined context. Straightforwardly, the proposition expressed by S is a structure in which 'I' gets replaced by a context-sensitive description.

The problem with this naïve view is that there is a large variety of observations showing that 'I' does not stand for a context-sensitive concept, for the very reason that it effectively acts as a directly referential term. It refers to a particular and contributes a proposition in which this particular is a direct component of the proposition itself. Consider for instance the sentence T = 'Well, I am here now,' uttered by D. Kaplan after his delayed arrival at a conference where he is the keynote speaker. Under the naïve view described above, the proposition expressed by T should be something along the lines of 'The utterer of T is at the place where he utters T at the time at which he utters T.' This is a tautology, that is, a proposition that is true in all possible worlds. In the context described above, however, Kaplan does not intend, of course, to express such a completely uninformative proposition. He clearly intends to express the proposition 'Well, Kaplan is finally at the site of the conference at time t.' This is informative in the relevant context (it causes for instance his audience to realize that the conference can start) and the reason is that the proposition is not true in all worlds: it might certainly be the case that Kaplan is not at the conference site at time t, and, in fact, a plausible reason why Kaplan uttered T was that there

was a serious risk for it to be the case. We should conclude that *T* stands for a proposition in which 'I' is replaced by the individual Kaplan, and is thus directly referential, not for a proposition in which 'I' is replaced by a context-sensitive description.

This much granted, there is certainly a sense in which the context is responsible for the reference of 'I.' Crucially, however, this context-sensitivity cannot be captured by proposing that the relevant proposition contains, in the place of 'I,' a context-sensitive concept. 'I' directly stands for the individual to which it refers *in a given context*. The concept is not part of the proposition, that is, the description is not part of the meaning expressed, though there still is a sense according to which the description contributes to determine which meaning is expressed.

Kaplan's solution to this puzzle has a technical flavor. He introduces two distinct levels of meaning, which correspond to duplicating the notion of 'world.' At the first level of meaning, dubbed *character*, meaning is relativized to the notion of world understood as the context, that is, the circumstances of evaluation of the utterance. The character is therefore a function from contexts to propositions. At the second level of meaning, dubbed *content*, meaning is relative to the standard notion of world, as we have understood it so far, that is, the situation the universe is in. The *content* is in fact a proposition, standardly conceived as a function from worlds to truth-values.

This 'two-dimensional' solution is sometimes regarded as 'baroque' (and rightly so), but there is a widespread consensus that these complications are unavoidable. The insight that the proposition expressed by a sentence containing 'I' directly features the individual referred to (rather than the context-sensitive concept by means of which the reference is established), captured by the duplication of meaning into character and content, undoubtedly has an explanatory power. Take the sentence *U* = 'I am D. Kaplan,' uttered by Kaplan while he is introducing himself at a party to someone who is not acquainted with him. Within a two-dimensional framework, the proposition expressed by *U* is 'D. Kaplan is D. Kaplan,' hardly something that does justice to the informational content proper to *U*. The problem is solved by taking the character of *U* into account: *U* is informative because, by uttering *U*, Kaplan provides information about the identity of the speaker. Similarly, consider again sentence *T* = 'I am here

now.' As we have seen, the proposition expressed by *T* directly features the referents of the indexical expressions in *T* ('Kaplan is at the conference site at time *t*'). The notion of content correctly captures the fact that this sentence does not express a necessary truth. However, the notion of character captures another, curious aspect of *T*: whereas the sentence may not express a necessary truth, it seems to express something that is true *a priori*. Ask yourself this: Is it possible to conceive of a context of utterance in which *T* is not true? The answer is 'no.' The sentence is bound to be made true by the very act of uttering it. Again, it is the necessary nature of character of the sentence that explains this property: How couldn't the speaker be at the place and time defined as the place and the time where the speaker is?

This two-dimensional apparatus is also quite useful in the analysis of the meaning effects produced by the other major class of directly referential expressions, such as proper names. Consider the Fregean sentence *Q* = 'Hesperus is Phosphorus.' As is well-known, both Hesperus and Phosphorus refer to planet Venus. Clearly, however, *Q* is informative. How can it be if both names are rigid designators, that is, if they both refer to Venus in all worlds *w*? The answer is arguably provided by the semantic contribution of the character of *Q*: the context in which the reference of Hesperus to Venus was established is different from the context in which the reference of Phosphorus to Venus was established. Realizing that *Q* holds was thus in no way a trivial epistemic effort, that is, Q is not true *a priori*.[10]

All in all, two-dimensional semantics may be 'baroque,' but it is certainly useful, especially because it captures the fundamental insight in which we are interested: indexicals and proper names are directly referential, in the sense that the individual entity to whom they refer to is a direct constituent of the proposition expressed.

---

10 Many scholars, among whom Kaplan himself, do not agree on the fact that this is a proper use of the two-dimensional framework originally developed by Kaplan, whose only goal was, after all, that of explaining the relation between the content of an indexical expression and its context of use, not that of accounting for the fact that sentences such as 'Hesperus is Phosphorus' express necessary, though not *a priori*, propositions (see Fiorin and Delfitto 2020, ch. 23).

Having acknowledged that, Kaplan's analysis is definitely not all we need to do justice to the direct referentiality of the first-person. This introduces, in fact, a quite fascinating topic, related to Perry's remarks about the involuntary sugar-spiller at a supermarket (Perry 1979). Suppose Perry thinks there is a sugar-spiller going around in the supermarket, since there is sugar everywhere on the floor. At a certain point, Perry realizes that he himself is the sugar-spiller. When this happens, he probably thinks: 'Uh, *I* am the sugar-spiller.' Perry's original observation is that the meaning expressed by the sentence above in the context described cannot be expressed by any other sentence. The reason is that this thought is essentially based on some primitive way in which Perry is given to himself.

In a recent contribution (Kripke 2011), S. Kripke explicitly raises this problem while discussing the explanatory power of Kaplan's account of indexicals, and relates Perry's observations to Frege's original preoccupations on the semantics of the first-person, which was heavily criticized by Kaplan and other two-dimensionalists. Here is the relevant quote from Frege:

> Now everyone is presented to himself in a special and primitive way, in which he is presented to no one else. So, when Dr Lauben has the thought that he was wounded, he will probably be basing it on this primitive way in which he is presented to himself. And only Dr Lauben himself can grasp thoughts specified in this way. (Frege 1918/19: 19)

Clearly, these thoughts are reminiscent of Descartes' insight that one is aware of himself in a special first-personal way. If one looks for a conceptual rendering of this primitive first-personal awareness, she immediately realizes that even complex concepts like 'the subject for me' or 'the subject I am aware of by being aware of my feeling and/or thinking' will not do, since all these formulations contain within themselves the notion of first-person, hence, as Kripke puts it, 'obviously run into a circle' (Kripke 2011: 298).

In order to identify the problem more precisely, it is useful to go back to Kaplan's formulation of the semantics of the first-person. We already know that 'I' cannot be replaced by a context-sensitive description. In fact, Kaplan's semantic rule for the interpretation of 'I' is given as a rule of *use*. Kripke invites us to imagine how such a rule would be formulated for German and how a German speaker (say, Frege) would then proceed to make use of such rule. Here is a plausible formulation of the rule:

> If any person *S* speaking German attributes a property using the word 'Ich,' then what *S* says or thinks is true if and only if *S* has that property.

Arguably, insurmountable problems arise. Here is how Kripke formulates them:

> But how can Frege use the word 'ich' on the basis of these instructions? Should he think, 'Hmm, so how am I going to use the word 'ich' on the basis of this general statement? Well, any German should attribute, say, being in pain or being a logician to himself if and only if the German is in pain or is a logician, as Kaplan says. So I should do this.' Alternatively, Frege might remark, 'So Frege, or Dr. Gustav Lauben, should attribute a property to Frege, or respectively to Dr. Lauben, using 'ich' if and only if Frege (or Dr. Lauben) has the property. But I am Frege, so I suppose that I should use the word 'ich' if and only if Frege has the property.' Either formulation would presuppose that Frege already has the concept of himself, the concept he expresses using 'ich,' so here we really are going in a circle. (Kripke 2011: 301)

Facts are thus more intriguing than in Kaplan's analysis. In fact, there are some extra ingredients to the puzzle of the first-person. The insight that 'I' refers to the special and primitive way in which everyone is given to herself is related to another exceptional feature of 'I,' which 'I' manifests in direct reports of mental states, as in Wittgenstein's famous example: 'I am in pain' (Wittgenstein 1958). A sentence such as 'I am in pain' is (in the terminology introduced by Shoemaker; Shoemaker 1968, 1996) *immune to error through misidentification* (IEM). This means that in uttering the sentence, one cannot possibly be wrong about the fact that who is in pain is not he himself but someone else. This happens only in particular situations. For example, suppose I utter the sentence in a context in which I am presented with pictures of twenty years ago, and that, by uttering it, I mean that the person I presently see in the picture (whom I identify with myself) is in pain. I might well be wrong: perhaps, the person in the picture is not me, but someone who strongly resembles me. This, however, cannot happen if the sentence is interpreted as a *bona fide* direct report on how I feel right now. In this scenario, misidentification is simply impossible.

This observation perfectly fits Kripke's (as well as Frege's original) point. IEM seems to be a consequence of the fact that the referent of 'I' correlates with the 'special and primitive way in which everyone is presented to himself.' Since this is a condition which is immediately and unreflectively given to us, there can be no issue of misidentification. In this case, reference is

really direct: there simply is no descriptive property to render the special and primitive way each of us is aware of herself.

This observation raises a number of serious semantic issues, which we will not further discuss here (but see Fiorin and Delfitto 2020). The conclusion to which we would like to draw the reader's attention is simply that the semantics of first-person constitutes a formidable case for direct reference.

Let us take stock. Arguably, there are sentences expressing propositions that are not exclusively made up of concepts. Certain linguistic expressions do not simply stand for concepts, but rather stand directly for their referents.[11] It is therefore not only the case that reference is rooted in the perceptual systems. It is arguably also the case that reference is sometimes established without any recourse to concepts. Or, if concepts are involved, they are arguably of a subjective, perspectival nature. This is particularly clear when we consider the essentially 'experiential' processes that underlie first-person reference, as discussed above. On the one hand, as some authors have suggested, pain – and more generally bodily sensations – can be considered a sense if there are dedicated transducers for it – this connects to the general issue of the 'individuation of the senses' (Matthen 2015b). On the other hand, if proprioception had to be categorized as a form of perception, shouldn't it be possible for pain – as well as for other bodily sensations – to be unconscious? In a sentence like 'I am in pain,' what is indeed the perceptual content of 'I'? Based on Frege's insight, 'I' should correspond in this case to the private 'experience' that everyone has of himself. Subtle and intriguing issues arise, since we may still ask what is the *distal object* involved in entertaining the proposition 'I am in pain.'[12] This is nothing less than Descartes' problem.[13] There may simply be no answer to this question, or we may regard it as one of the typical manifestations of the so-called 'hard-problem of consciousness.[14]

---

11 Whatever these referents exactly are; this issue, of course, becomes especially critical in the case of 'I' when used in IEM contexts.
12 On this see Shoemaker (1994).
13 See Anscombe (1975) and the discussion in Fiorin and Delfitto (2020, ch. 29).
14 See Chalmers (1996).

3.4 Direct reference in the visual system

So far, we have reached two main conclusions:

i. Reference is originally established in acts of perception;
ii. Propositions are not internal properties of the mind, since they are not made solely of concepts.

Regarding (ii), one might still contend that the fact that certain linguistic expressions (such as proper names and indexicals) refer directly reveals something of the complexity of the systems of language and thought, as distinct from perception. Whatever the source, nature, and workings of the system of *direct reference*, it is a property of language as a higher-order cognitive system. Now, this claim is simply not correct. In fact, it is a patent denial of how things actually are. Direct reference is one of the most prominent and fundamental properties of the visual system. It is useful to consider this point in some detail.

It was long believed that the way we refer to objects in vision has nothing to do with direct reference. Objects are tracked – it was generally thought – by binding a set of properties at one and the same location and are distinguished one from another by associating them with different sets of properties. A large bulk of experimental work by Kahneman and, especially, Pylyshyn and his associates, has revealed that this picture is not empirically warranted. To consider a notable example, Kahneman and Treisman (1992) demonstrates that letter priming travels with the object in which the prime first appeared. In this experiment subjects see letters in boxes that then are cleared and moved to a new location. Then a letter appears in one of the boxes, either the same box it had been in or a different box. Subjects must then name the letter. When the letter reappears in the same box, it is named more quickly than if it appears in the other box, even after properties that could have favored that box (such as distance from the priming event) have been controlled.

Based on this and similar results, Pylyshyn explicitly contends that vision tracks objects by means of a mechanism of *direct reference*. These are the so-called FINSTs, that is, 'fingers of instantiation.' Here is a relevant fragment from Pylyshyn:

I have proposed that the capacity to individuate and track several independently moving things is accomplished by a mechanism in the early vision module that I have called FINSTs (I call them 'Fingers of INSTantiation' because they were initially viewed as a mechanism for instantiating or binding the arguments of visual predicates to objects in the world). This primitive non-conceptual mechanism functions to identify, re-identify, and track distal objects. It is an ability that we exercise every waking minute, and it has also been understood to be fundamental to the way we see and understand the world. (Pylyshyn 2007: Preface, x)

What this mechanism actually entails is that we see things primarily as 'unidentified' objects, objects for which it is cognitively irrelevant with which properties they are endowed. Seeing is, in other words, primarily non-conceptual: [FINSTs] serve to provide what philosophers have called demonstrative reference or demonstrative identification. […] FINSTs serve, namely, the non-conceptual tracking of individual things that move and change their properties. Since tracking is one of the critical aspects of our commerce with the world [the experiments we will discuss] serve as concrete examples of the role that FINSTs play in this process. It also provides a basis for a number of additional properties of this mind–world connection: it shows that things can be tracked as unidentified things with an enduring numerical identity (where by 'unidentified' I mean they are not represented in terms of any conceptual category or in terms of distinctive properties). (Pylyshyn 2007: Preface, xi)

Whereas much of the psychological and philosophical literature sees the binding problem as being solved in terms of the collocation of properties, my proposal is that properties are considered conjoined if they are properties of the same FINSTed thing. (Pylyshyn 2007: Preface, xii)

FINSTs have nothing to do with playing around with concepts and using concepts in abstract computations. Rather, FINSTs are inherently related to space and movement:

When my colleagues and I first came across this problem in the context of incrementally constructing a representation of a geometrical diagram it seemed to us that what we needed is something like an elastic finger: a finger that could be placed on salient things in a scene so we could keep track of them as being the same token individuals while we constructed the representation, including when we moved the direction of gaze or the focus of attention.' (Pylyshyn 2007: 13)

What is particularly striking, in our opinion, is that when it comes to illustrating how object-tracking actually works in terms of FINSTs, Pylyshyn uses *linguistic* processes of direct reference as a sort of exemplification, or if you want 'metaphor,' of what takes place in visual perception. One of the properties that define direct reference in language is in fact the core

property of reference in vision. It is worth going through some fairly long quotes. Here is how visual reference does *not* work:

> In the example sketched earlier, where we are constructing a description of a figure over time, we need to keep track of individual things so as to be able to determine which is which over time—that is, we need to be able to decide between 'there it is again' and 'here is a new one.' We must be able to do this in order to put new information into correspondence with the right individuals already stored in memory. We also need to be able to decide when we have noticed a new individual thing or merely re-noticed one we had already encoded earlier. Being able to place individual things into correspondence over time—or to keep track of individual tokens—is essential to the ability to construct a coherent representation. When we notice an individual thing with property P we must attribute P to the existing representation of that very token (if we had encoded it before), or else we must augment our stored representation to include a new individual thing. One way to place individual things into correspondence is to associate a particular token thing with what Bertrand Russell called a definite description, such as 'the object x that has property P' where P uniquely picks out a particular thing. In that case, in order to add new information, such as that this particular thing also has property Q, one would add the new predicate 'Q' to the representation of that very thing. This way of adding information would require adding a new predicate 'Q' to the representation of an object that is picked out by a certain descriptor. To do that would require first recalling the description under which x was last encoded and then conjoining to it the new descriptor. Each time an object was encountered once again, we would somehow have to find the description under which that same object had been encoded earlier. (Pylyshyn 2007: 14–15)

In a nutshell, if tracking particulars consisted in playing around with descriptions (assigning them to the object, resuming them when the object is seen again, etc.), object-tracking would be computationally intractable. Here is instead how object-tracking actually works:

> The alternative to this unwieldy method is to allow the descriptive apparatus to make use of the equivalent of singular terms or names or demonstratives. If we do that, then adding new information would amount to adding the predicate Q(a) to the representation of a particular thing a, and so on for each newly noticed property of a. Empirical evidence […] suggests that the visual system's Q-detector recognizes instances of the property Q as a property of a particular visible object, such as object a. This is the most natural way to view the introduction of new visual properties to the sensorium. This view is consonant with considerable evidence that has been marshaled in favor of what is referred to as object-based attention […]. In order to introduce new properties in that way, however, there would have to be a nondescriptive way of picking out a. This is, in effect, what the labels on objects in a diagram are for and what demonstrative terms like 'this'

or 'that' allow one to do in natural language. So what I am in effect proposing
is that the visual system needs such a mechanism of demonstratives. The object-
based view of how properties of objects are detected and encoded would sug-
gest that when we detect a new property we detect it as applying to a particular
object, rather than as applying to any object that has a certain (recalled) property.
(Pylyshyn 2007: 16)

And here is what the whole story essentially amounts to:[15]

> The empirical part of this story is the hypothesis that what perception initially
> detects is things or objects, as opposed to properties or locations. The more gen-
> eral claim, that something in the world is detected without prior specification of
> its properties, is more than an empirical hypothesis. (Pylyshyn 2007: 17)

In fact, this is still not the whole story. The essentially deictic nature of ref-
erence to objects in vision is not limited to the process of object-tracking.
As is well-known, certain aspects of vision are analyzed in parallel by dis-
tinct pathways, one of which is involved in object recognition and another
in visually guided movements. More specifically, the primary visual cortex
constitutes the first level of cortical processing of visual information. From
there, as already hinted at in Chapter 2, information is transmitted over
two major pathways: a ventral pathway into the temporal lobe carries
information about what the stimulus is, whereas a dorsal pathway into the
parietal lobe carries information about where the stimulus is, information
that is critical for guiding movement. Now, there is substantial evidence to
the effect that this motion-guiding information is essentially non-conceptual
in nature. It involves direct/demonstrative reference and abstracts away
from object-identification and description. The presence of an object in
the deictic space is immediately given. Here is how Matthen reports on
experimental evidence suggesting that motor-guiding vision is unconscious
vision, completely indifferent to descriptive content as a potential vehicle
for object-identification:

> In one experiment, subjects were told that a small light would suddenly appear in
> the periphery of their visual fields, and instructed to reach for it when it did. While
> they were reaching for the light, it was displaced. The light was moved during an
> eye-saccade: since they were looking in a different direction, the subjects did not

---

15 See Pylyshyn 2007 and Fodor and Pylyshyn 2015 for a full discussion of all
   relevant issues.

> see it move. When they looked again, the light was at a new position. Subjects failed to notice the shift, but were able to reach for the shifting light smoothly (Goodale, Pélisson, and Prablanc 1986). (Matthen 2005: 299)

We have suggested that conscious vision (in the sense of vision that is associated with the 'feeling' or 'experience' of seeing) involves the decomposition of a complex visual scene into a set of propositional events of attribution of properties to particulars. Seeing is thus a complex set of acts of predication. We also contended that each of these cognitive events of predication is potentially veridical: they are veridical whenever the conditions for correct seeing are satisfied (say, the laws of optics for the relation between distal and proximal object and the laws of formation for the relation between proximal object and final representational state). This attracting picture has a potential gap. It seems that when a cognitive agent is confronted with a complex visual scene, he is induced to entertain a rich set of propositions. However, in language propositions are not simply *entertained*, they can also be *asserted*, whereby a commitment arises, on the part of the cognitive agent, for the veridicality of the proposition that is entertained. One may then wonder whether *assertion* has a counterpart in perception. Matthen proposes that deictic seeing, as carried out, according to him, in the dorsal stream, can perform exactly this function, potentially filling this potential gap between language and perception. Though we are not ready to endorse this specific point, here is an extended quote:

> Our visual states present us with an assembled message, a message that has a descriptive element as well as a referential one. Motion-guiding vision is responsible for the latter. This referential element of visual states constitutes a kind of direct connection between perceiver and distal stimulus, and creates a feeling of reality of presence. Consider a singular proposition: 'John is tall.' This proposition can be entertained without being asserted. Similarly, a visual scene can be imagined or dreamed. In normal visual perception, however, the scene is not simply imaged, but seems to present the perceiver's own surroundings as so. I will refer to this as a 'feeling of presence.' The feeling of presence is similar to assertion: attached to a visual scene, the feeling of presence asserts it. (Matthen 2005: 305)

Matthen's deictic seeing, as well as Pylyshyn's non-conceptual FINST dynamics in object-tracking, have interesting epistemic properties. Though there is no argument to the effect that they are endowed with the sort of 'logical' immunity to error through misidentification that has been detected

in certain uses of the first-person,[16] they exhibit a sort of *de facto* immunity to certain optical illusions to which descriptive vision is easily prone. This is at least what is suggested by the experimental evidence around the three-dimensional variant of the Titchener's Circle Illusion:

> The Titchener circles illusion consists of two equal-size circles, one surrounded by an annulus of larger circles, the other by one of smaller circles. The circle surrounded by smaller circles looks bigger; in order to make the two circles look equal, this circle needs to be reduced in size. This experiment consisted in recreating the Titchener circles illusion with flat poker-chip type discs positioned in front of a normally sighted observer. All observers saw the central discs in the illusory way, but in reaching for them, their grip was scaled according to the actual size of the disc (Aglioti, DeSousa, and Goodale 1995). (Matthen 2005: 310)

In this sense, unconscious deictic vision improves epistemically on descriptive conceptual vision, an unexpected finding, especially when one considers that it is the latter, not the former, that gives rise to the experiential *qualia* associated to the descriptive features that we most typically take as defining for 'seeing':

> A (normally sighted) person might be able to reach for an object, but not see where it is: subjects were able to reach for something that had shifted during an eye-saccade though they did not consciously notice the change of location. This indicates that subjects are able to establish a sensorimotor connection with an object, to become 'perceptually coupled' with it as Noë and O'Regan say (2002 , 571), without being conscious of its location […] The feeling of presence that accompanies seeing real objects, which I have been connecting with the ability to reach out and touch or manipulate these objects, seems to be, as far as conscious visual experience is concerned, a pure demonstrative, a cognitive relationship between perceiver and object, devoid of all descriptive content. (Matthen 2005: 319)

Matthen's conclusion is that reference in perception is modeled by a complex dynamics between descriptive space (*D-space*) and deictic space (*N-space*), which is highly reminiscent of the two-dimensionality that seems

---

16 For *de se* readings and immunity to error through misidentification, see Wittgenstein 1958, Castañeda 1968, Anscombe 1975, Higginbotham 2003, Recanati 2007/2012, García-Carpintero 2015, Capone 2016, Delfitto, Fiorin and Reboul 2017, Fiorin and Delfitto 2020; for a syntactic perspective, see Tenny 2006; for a perspective on the topic from cognitive neuroscience, see Gallagher 2000.

appropriate for an adequate description of the properties of reference in language, as a viable attempt at reconciling the Fregean view with Kripke's and Kaplan's insights:

> The space in which we see objects is charted by a combination of motion-guiding and descriptive devices. (Matthen 2005: 322)

Ultimately, Matthen's view is tantamount to contending that objects (direct reference) are epistemically prior to properties (descriptive reference):

> In the more standard presentation, features have an independent status; material objects are constructed out of them. In the alternative version presented here, material objects come first; features are attributed to them after they are identified […] After motion-guiding vision identifies a material object as something to attend to, descriptive vision assigns it to descriptive classes (Matthen 2005: 324)

We do not share Matthen's view on this point. The way he refers to direct reference in vision as essentially linked to vision-for-action downplays Pylyshyn's arguments in favor of a non-descriptive view of vision-for-cognition. In fact, FINSTs show that there is direct reference in vision even abstracting away from action-directed visual stimuli.[17] Non-descriptive object-tracking, in Pylyshyn's sense, is not linked to action *per se*. In a series of important contributions, Milner and Goodale (see Goodale and Milner 1996, Milner and Goodale 2006) proposed that the ventral and dorsal visual systems, which evolved for object perception on one hand and for action-control as directed to those objects on the other hand, are uniformly modulated by attention and uniformly based on information processing concerning the structure of the objects and their location in space. However, it is only the ventral system that transforms the visual input into representations corresponding to the enduring features of objects and their relations, whereas the execution of object-directed actions is carried out by dedicated control-systems in the dorsal stream. It is uncontroversial – we believe – that visual direct reference extends to both the dorsal and the ventral stream, a conclusion supported by the fact that there is a substantial amount of interaction between the dorsal and the ventral pathways. This is not to deny that the way of operating of the action-oriented dorsal stream is essentially non-descriptive, from the use of simple visual information

---

17 We are indebted to Anne Reboul for drawing our attention to this point.

about simplified shape, width and orientation (rather than about fully-fledged objects), to the absence of a delay between visual stimulus and motor response, as well as the absence of a memory of the stimulus, to the incapacity of reporting manually shape and orientation of the object after the stimulus has gone. In other words, though direct reference cuts across the two visual streams, action-directed vision in the dorsal stream is inherently non-descriptive and supports the core role played by deictic reference in vision as a whole. Crucially, however, this is not the same as proposing, as Matthen does, that there is a sort of division of labor, concerning deictic vs. descriptive reference, between the ventral and the dorsal pathway. To put it shortly, there is no empirical foundation for the claim that direct reference is an exclusive prerogative of action-oriented vision as encoded in the dorsal stream.

Putting these important qualifications aside, there cannot be doubts that the conception of reference that we have laid out above is orthogonally opposed to Hinzen's proposal that reference arises with grammar and is, in fact, built up from grammar:

> I will call the specific kind of meaning that goes with grammatical organization grammatical meaning in what follows. With grammar, we can refer and predicate, and the result is propositional truth. Along with truth comes another distinctive human privilege: the making of mistakes. (Hinzen 2014: 235)
>
> Adding a grammatical system to a system of concepts does not then result in more concepts, but instead something entirely different: a capacity for using a resource of stored concepts (semantic memory) for a new purpose, namely referentiality, which does not exist in the same form in animals (Fitch 2010). That is:
>
> (1) a. WRONG EQUATION: concepts + grammar = more concepts
>   b. RIGHT EQUATION: concept + grammar = reference.
>   (Hinzen 2014: 235)
>
> Consider a simple progression like the following, which has nothing to do with a change in the conceptual content of the lexical item dog, but instead solely with different ways in which it can be embedded in the deictic space that the speaker shares with the hearer:
>
> (2) I ate **dog** < I ate **dogs** < I ate **a dog** < I ate **the dog** < I ate **this dog**
>
> […] We therefore move from a maximally indefinite to a maximally definite form of reference, as grammatical complexity of the determiner phrase increases. This process finds an end when the deictic stage is reached and referential specificity is maximal. (Hinzen 2014: 238)

These quotes from Hinzen express in effect two distinct insights, the first of which is seriously flawed and the second of which is probably correct.

The wrong insight is that reference is a creation of grammar. We have argued not only that reference is an original property of perception, we have also demonstrated that one of the most distinctive and controversial properties of reference in language, that is, non-conceptual direct reference, is deeply rooted in vision and effectively contributes to explaining how particulars are identified, how property-ascription actually takes place, and how vision feeds object-directed action. The correct insight is that language creates many further distinct ways of referring: for instance, reference to mass and countable entities is quite different in English and Chinese[18] and the tricks of coreference are deeply rooted in language design.[19] This is part of a more general phenomenon already hinted at above, consisting in the fact that language plays around with perceptual content and creatively acts on it by modifying and extending it (this is what happens, for instance, with color terms in language). More precisely, this is part of the observation we made above, to the effect that the concept repository considerably increases once the conceptual atoms are put within the universally defined 'logical' space associated with the functional architecture of language. Indeed, these refinements and extensions of the ways of referring in language arguably take place by resorting to the special properties of the functional architecture in syntax.[20]

Significantly, Hinzen seems to regard the *deictic stage* as the maximally specified aspect of reference or, if you wish, the ultimate product of the creative power of grammar. Again, we should emphasize that this insight is flawed. Direct reference is a creation of visual perception, and, in many ways, it would be more correct to go for the opposite claim that syntax, by encoding direct reference in the formal templates of the functional lexicon, simply *mirrors* the representational power of the visual system.

---

18 Chierchia (1997); Cheng & Sybesma (2012).

19 Reuland (2011).

20 See especially Borer's (2013) exoskeletal approach to words and syntactic categories. For a highly influential view of how syntax affects reference in language, see Longobardi (1994).

From this viewpoint, the most challenging question is then the following: *Where does this mirroring come from?* Or, from a slightly different and perhaps more puzzling perspective, how come that some formal patterns of syntax (of which we are entirely unconscious) somehow duplicate the abstract properties of visual perception (like non-conceptual reference), of which we are also completely unaware, as witnessed by the fact that Pylyshyn's modeling of object-identification in vision, based on the primacy of the object over the properties that define its constitution, remains largely counter-intuitive, in spite of the substantial amount of experimental evidence supporting it?

Direct reference is most certainly *not* a creation of grammar. And direct reference in grammar is certainly *not* the creation of conscious acts of reflection on some abstract properties of perception. On this basis, it rather looks like direct reference provides a case for the thesis that both vision and grammar are permeated by the same abstract conditions for objective representation. We will refer to this view as the *Harmony Thesis*.

Certainly, one might counter that one swallow does not make a summer. That may be so, but in the next two chapters, we will see that there is more to the Harmony Thesis than just direct reference.

Chapter 4   What does grammar tell us about seeing?

4.1 Syntax and ways of seeing

The syntactic literature makes us acquainted with the fact that the verb of perception *'see'* may select, in English, two different sentential syntactic complements. The first is the that-clause (shown in (1)), the second is the so-called *naked infinitival* (shown in (2)):

(1) Bill saw [that Liz left by bike]
(2) Bill saw [Liz leave by bike]

These two constructions manifest two complex sets of distinct interpretive properties, which clearly keep them apart from each other. Taken as a whole, these two distinct sets of properties seem to argue for the conclusion that (1) expresses *epistemic* seeing and (2) expresses *perceptual* seeing. We will see that this terminology can be misleading, especially for what concerns alleged *perceptual* seeing. But let us first consider what the properties of (1) and (2) are.

   Suppose Bill sees every morning his neighbor Liz leave for school by bike. Yesterday, however, something unexpected happened: Liz did not show up at all, and Bill was witness of that. Suppose further we are trying to reconstruct what caused Liz not to go to school that morning. While doing that, we might conveniently report what happened by uttering (3) but not, crucially, by uttering (4):

(3) Bill saw that Liz didn't leave by bike
(4) Bill saw Liz not leave by bike

Interestingly, something remarkable happens when we displace negation from the embedded clause (the that-clause and the naked infinitival) to the main clause, giving rise to (5) and (6), respectively. In this case, it is (5) that no longer is an adequate report of the relevant situation, while (6) is perfectly fine:

(5) Bill didn't see that Liz left by bike
(6) Bill didn't see Liz leave by bike

What is going on? Upon reflection, we conclude that (5) will not do because by uttering it we are presupposing that Liz left, and we are asserting that for some reason Bill did not have visual experiences related to her leaving. A more formal way to describe this interpretation is to say that (5) entails that Liz's leaving is a *fact*, and that Bill did not have a visually justified knowledge of this fact. However, in the situation described above, we do not know whether Liz left or not, and this explains why (5) is an inappropriate report. Conversely, the intuition we have about (6) is that by uttering it we are simply denying that Bill had a visual experience of Liz's leaving, without any commitment as to whether Liz left or not. This is arguably a correct report of the situation. Even here, a perhaps better way to convey these remarks consists in saying that (5) is about *visually justified knowledge of a fact*, whereas (6) is about *visual experiences*. More particularly, (5) denies that Bill has a visually justified knowledge of the fact that Liz left (an inappropriate description of the relevant situation), whilst (6) denies that Bill had a visual experience whose content is Liz's leaving, which is an appropriate description of the relevant situation, it seems.

This line of analysis is confirmed by the contrast between the negative sentences (3) and (4), which supports the generalizations introduced for (5) and (6). By uttering (3), we state that Bill knows that Liz did not leave, based on what he saw. Interestingly, in fact, what he saw is that Liz did not show up, the bike remained in the garage, etc., and this perceptual content seems to be enough to support the epistemic conclusion that Liz did not leave. In a nutshell: that Liz didn't leave is a fact, and Bill has some visually grounded knowledge of this fact. Or even better: whoever utters (3) presupposes that Liz did not leave as a fact, and asserts that Bill's knowledge of this fact is based on the perceptual contents stemming from Bill's perceptual experience. Conversely, by uttering (4) we cannot possibly intend that Bill visually experienced a negative event, since what the perceptual experience of a negative event should amount to is virtually unintelligible. We may know that a certain fact did not occur, but this must be a matter of inference from the visual knowledge of some positive facts: there is no direct perceptual experience of a negative event. The way out consists in assigning a narrow scope to negation. The latter may take scope over the VP, in which case what we are saying is that Bill visually witnessed something Liz did, but *that* something was not a leaving by

bike. This reading becomes promptly available in pragmatically supportive contexts. Alternatively, by exploiting the logical forms associated to standard event semantics (Parsons 1972), negation is assigned narrow scope over the instrumental argument: essentially, we are asserting that there was a leaving by Liz, but we are denying that this leaving was by bike.

This analysis of the intriguing behavior of negation in (3–6) has important consequences. It entails that the use of a different syntactic complement in (1–2) deeply affects the semantics of '*see*.' In (1) the that-clause refers to a *fact* and 'see' is interpreted *epistemically*, that is, as involving visually justified belief that something is the case. In epistemic seeing, the relationship between justified belief and (visual) perceptual experience is not necessarily direct. As is well-known (Jacob and Jeannerod 2003), one could state that she saw 'that her dog crossed the road' in a situation in which what she actually saw was her dog, on one side of the road at time *t*, and her dog on the other side of road at a time *t'* following *t*. These two virtually independent visual experiences may be the grounding perceptual experiences for one to know (by inference) that a certain fact holds (the dog crossed the road). In other words, what is expressed by the that-clause complement of 'see' is not really perceptual content. It is rather the propositional content that is the result of the interaction of one's cognitive systems (one knows, for instance, about the causal relations between two visually witnessed events) and one's perceptual systems (what one directly experiences in perception). Crucially, in the situation described above, one could not state that she saw 'her dog cross the road.' This sentence entails that she had a direct visual experience of her dog crossing the road, while this is not the case in the situation at stake.

In what follows, we will further examine what the divide is between epistemic and perceptual seeing, as well as some consequences of this divide. For the time being, notice that an important conclusion is already warranted: the contrast between different ways of seeing is expressed, in language, by means of dedicated syntactic templates (that-clauses vs. naked infinitivals).

Consider now the following property of epistemic seeing. When a that-clause is the complement of a verb of propositional attitude, such as 'believe,' it typically gives rise to *de re* / *de dicto* ambiguities. This can be seen in the interpretation of (7):

(7) Bill believes that the president of the US is incompetent

A *de re* interpretation of (7) is an interpretation according to which there is an individual that is the president of the US (say, Joe Biden) and Bill believes of *that* individual that he is incompetent. A *de dicto* interpretation of (7) is an interpretation according to which Bill believes that a person he deems to be the president of the US is incompetent, whereby the identity between *that* person and the actual president of the US (say, Joe Biden) is not ensured. The question we are interested in is whether epistemic seeing gives rise to this kind of ambiguity. The answer is a loud 'no.'

To see this in some detail, consider the example of epistemic seeing in (8):

(8) Bill saw that a dog crossed the road

A sentence such as (8) cannot be continued as in (9):

(9) ??Bill saw that a dog crossed the road, and he never realized it was actually a cat

Suppose that the indefinite noun-phrase 'a dog' is interpreted *de dicto* in (8). This means that Bill saw an animal *as a dog* and that he saw that that animal crossed the road. If this is the interpretation of (8), we predict that (9) should be fine. In principle, there is nothing wrong with pointing out that Bill was in fact the victim of an optical illusion concerning the animal that crossed the road. However, the trouble is that whoever utters (9) is bound to deal with the content of the that-clause in (9) as a fact, that is, she is bound to presuppose the truth of the proposition according to which a dog crossed the street. It follows that what is asserted in the continuation (that is, Bill's coming to realize that the animal he had seen *as a dog* was actually a cat) provides an utter contradiction: for (9) to be true it should be the case that the animal that crossed the road was both a dog and a cat. This explain why (9) sounds weird and hardly acceptable. It also demonstrates that the indefinite noun-phrase 'a dog' cannot be interpreted *de dicto*. The property according to which the animal crossing the road is seen as a dog[21] is compulsorily relativized to the speaker's cognitive/perceptual system, and cannot thus be relativized to Bill's cognitive/perceptual system. This boils

---

21 The careful reader has probably already realized that 'see as' is used here in the sense of Dretske (1981; 2000).

down to concluding that 'a dog' can only be interpreted *de re* in contexts of epistemic seeing, a conclusion that squares with the traditionally acknowledged *factive* status of the that-clause complement of 'see.'

As discussed above, the complement of epistemic 'see' is not a perceived event or a visual experience, but a *fact* of which the reporting agent knows, on visual grounds, that it holds. If this is the case, we predict that the ways of describing the object referred to in the complement of 'see' necessarily reflect the perspective of the reporting agent, that is, of the utterer of the sentence.

Remember that, in this respect, in Chapter 1, we had examples like (10) below:

(10) ??Yesterday, I saw that my baby was smiling, but in fact it was not my baby

In principle, one might be surprised that (10) could not be used as a report of an optical illusion I had yesterday: as such a report, (10) would simply state that it was someone else's baby who was smiling. However, the observations above directly apply. In asserting that I saw that my baby was smiling I am necessarily presupposing the truth of the proposition that my baby smiled, and this truth cannot be denied in the continuation of the sentence.

What holds of optical illusions also holds of hallucinations. Consider (11):

(11) ??Yesterday, I saw that a baby was smiling, but actually there was no baby around

Again, the that-clause refers to a fact, and denying that this fact holds results in a contradiction. This explains the oddness of (11) as a potential hallucination report.

Other observations pledge for a *factive* analysis of epistemic seeing. Though we will not discuss them in detail here, let us briefly mention them. First, the that-clause complement of *'see'* behaves, in English, as a 'weak island' for extraction purposes, exactly on a par with the behavior of uncontroversial *factive* predicates such as 'regret.' Extraction of an adjunct from the that-clause in (12a) yields a deviant result (12b).

(12) a. I saw that Liz left by bike
     b. *?How did you see that Liz left?

Second, Saito (2015) convincingly shows that complements of perception and factive verbs have identical syntactic forms in Japanese, especially for what concerns the distribution of the two complementizers 'to' (found with verbs of propositional attitudes) and 'no' (found with factive verbs and verbs of perception), and further argues that the differences between the complements of factive verbs and of verbs of perception are due to independent factors. Third, potential counterexamples to the claim that the factive presupposition proper to the complement of perception verbs cannot be contradicted in the continuation of the relevant sentence are easily dismissed by the recourse to the role played by some independent additional factor. A case in point is (13), where the factivity of the complement is cast in doubt in the second part of the sentence:

(13) Bill could not see that his wife left the building, for the very reason
       that she never was inside the building

Here, it is easy to see that these cases amount to a standard use of *metalinguistic negation*, whereby the latter denies the factive presupposition conveyed by (13), that is, the proposition that Bill's wife left the building (Moeschler 2014).

4.2 Knowing facts (by seeing) and seeing events

Knowledge is canonically defined as follows:[22]

(14) *a* knows that *p* if and only if:

    (i) *p* is true;
   (ii) *a* believes *p*;
  (iii) *a*'s belief is justified.

A reasonable interpretation of (14iii) takes it that one of the ways in which belief can be justified is by means of reliable perceptual processes. As discussed in Chapter 2, perceptual acts are endowed with representational/propositional content, and this content may easily lead to true

---

22 This tripartite definition of knowledge as 'justified true belief' is actually not
     adequate, since one can easily build up many cases of justified belief that still
     fall short of knowledge. The interested reader is specifically referred to what
     came to be known as the 'Gettier cases' (Gettier 1963).

belief, characterized as, say, the disposition to assent to the propositions that express that very perceptual content. The final belief state need not coincide with the propositional content associated with the actual acts of perception, since the former can result from cognitive penetration, that is, be produced by the application of inferential devices that are part of the cognitive endowment of the perceptual agent.

On these grounds, *epistemic seeing* can be defined as follows:

(15) *a* sees that *p* if and only if

  (i) *a* knows that *p*;
  (ii) *a*'s knowledge is justified by *perceptually reliable* visual processes.

This analysis automatically delivers the desired loose relation between the perceptual events of vision that justify the knowledge that *p* and the final propositional content of that knowledge, that is, the content of *p*. If one says that she saw that Liz left, the knowledge that Liz left need not be justified by visual witnessing Liz's leaving. One may have, for instance, the visual experience of Liz's house as empty, or the visual experience of Liz's bike no longer being at its usual place. As repeatedly emphasized, these two visual experiences may be deemed as sufficient to 'infer' that Liz left.

Epistemic seeing is, in a sense, 'theoretical' seeing. A hardly surprising finding, since we know that 'theory' is etymologically related to the Greek verb '*theōréō,*' that is, seeing. Vision typically feeds knowledge, as also witnessed by the common extension of the epistemic interpretation of 'see,' in many languages, to contexts in which 'see' simply means 'know,' without any requirement that the relevant knowledge be visually justified, as when we say, in English, that we 'see' that someone is right or that we 'see' that a certain theorem has been proved. Again, a hardly surprising finding, especially when considered from the perspective we are adopting in this book, according to which vision (and perception in general) is an inherently propositional cognitive capacity.

Consider now some of the properties of what we have dubbed *'perceptual seeing.'* The sentences in (16) are both fine:

(16) a. Yesterday I saw my baby smile, but actually it was not my baby
     b. Yesterday I saw my baby smile, but actually there was no baby around

When the complement of 'see' is a naked infinitival, it can be used to express the content of an optical illusion, as is the case in (16a), or the content of a hallucination, as is the case in (16b). Here, the questions that we should ask are essentially two:

 i. Which are the properties of the complement of perceptual 'see'?
 ii. Which kind of objects does perceptual 'see' refer to?

These two questions are strictly intertwined. The syntactic literature has made us acquainted with three major properties of the naked infinitival complement of 'see': (i) it is incompatible with modals; (ii) it is incompatible with sentential adverbs; (iii) it exhibits a peculiar behavior with sentential negation (see Bayer 1986, Barwise 1982).

The incompatibility with modals is exemplified by the German sentences in (17) (from Bayer 1986):

(17) a. *Wir sahen Hans tanzen können
          we saw Hans dance can
      b. *Wir sahen Hans singen müssen
          we saw Hans sing must
      c. *Wir sahen Hans ins Kino gehen wollen
          we saw Hans to-the movie go want

The incompatibility with sentential adverbs is shown by the severe deviance of the English sentences in (18):

(18) a. *He saw John certainly murder the woman
      b. *He saw John clearly murder the woman
      c. *He saw John without doubt murder the woman

Notice that the sentential adverbs in (18) yield a fully acceptable grammatical result when used in the German counterparts of (18), as shown in (19) below:

(19) a. Er sah den Hans sicherlich die Frau ermorden
      b. Er sah den Hans ganz klar die Frau ermorden
      c. Er sah den Hans zweifellos die Frau ermorden

Significantly, however, the relevant adverbs can only be interpreted, in (19), as modifying the matrix clause. They cannot be interpreted as modifying the perceptual content expressed by the embedded clause.

It goes without saying that these data invite an interpretation of the complement of perceptual 'see' as referring to *events* that can be perceived, not to facts/propositions. What one can perceive – or so the argument goes – are objects and events, not facts. The syntactic constraints illustrated above (incompatibility with modals and sentential adverbs) are easy to reconcile with the idea that the infinitival complement corresponds to a reduced clause-like syntactic constituent, which essentially instantiates the templatic structure that assembles the lexical items that encode the conceptually salient event properties, while completely abstracting away from temporal and world parameters. Semantically, and quite interestingly from the present perspective, this insight somehow implements the essentialist position (Fine 2005) according to which there is a level of event conceptualization that projects into atemporal structures and that we might be tempted to identify with the basic ingredients of perceptual content, that is, with the part of propositional content that is shared by language and perception (see especially Ramchand 2018).

The observations on negation that we have made above point in the same direction. First of all, there is a complex relation between the logical form in which negation is in the main clause and the logical form in which negation is in the complement clause. The conditional in (20) seems to hold (see Barwise 1982):

(20) If a sees NOT φ, then NOT (a sees φ)

This translates in valid inferences of the form 'If Bill saw Liz not leave, then he didn't see Liz leave.'

Second, as we have seen above, the logical form 'a sees NOT φ' demands partial scope of the negation within the embedded clause, on metaphysical grounds. If the complement of 'see' is bound to refer to an event that can be perceived, there is no room, metaphysically, for negative events (how could a negative event be perceived?). It follows, as we have proposed above, that a sentence like 'Bill saw Liz not leave' should be interpreted as the report on a perceptual event in which Bill saw Liz do something, though this something was not leaving.

French offers a direct confirmation that this 'narrow scope' analysis of negation is correct. Consider the two sentences in (21):

(21) a. J'ai vu que tous les etudiants n'ont pas parlé au professeur
          I have seen that all the students not have spoken to the professor
      b. J'ai vu tous les etudiants ne pas parler au professeur
          I have seen all the students not speak to the professor

As is well-know, sentential negation can be construed ambiguously in sentences such as (21a), where it interacts with a universal quantifier in subject position. Under the narrow scope construal, (21a) means that I saw that all of the students didn't speak to the professor (that is, no student spoke to the professor), whereas under the wide scope construal (21b) means that I saw that not all the students spoke to the professor (i.e. some students did, some other students did not). However, when the complement is a *'naked infinitival,'* as in (21b), only the narrow scope construal is admissible. The wide scope reading, according to which I saw not all the students speak to the professor, is logically conceivable but grammatically not legitimate, that is, it does not constitute a viable interpretation of (21b) in French. This is exactly what the analysis proposed above predicts: negation should not be able to get scope over the whole clause, since that would turn the event referred to into a negative event, and a negative event cannot be the object of acts of visual perception.

The hypothesis that the reference of the infinitival complement of 'see' is an event is further confirmed, arguably, by the behavior of Tense in the embedded clause. In the case of the that-clauses complement of epistemic 'see,' there is no tense constraint, as shown by the full acceptability of both sentences in (22):

(22) a. Bill sees that Liz is leaving
      b. Bill sees that Liz left

On the contrary, the past tense is inadmissible in the infinitival complement of *perceptual 'see,'* as shown by the contrast in acceptability between (23a) and (23b):

(23) a. Bill sees Liz leave
      b. *Bill sees Liz have left

Arguably, this constraint simply reflects a physical condition on correct perception: the perceived event and the act of perception must be co-temporal.

Moreover, the sharp contrast between epistemic and perceptual seeing clearly emerges as a contrast in acceptability between (24a) and (24b) below, under the folk-psychological assumption that dogs can be granted the possession of the capacity of visually perceiving objects and events but not the possession of visually justified knowledge of facts (Bayer 1986).

(24) a. The dog saw Bill steal the money
      b. ??The dog saw that Bill stole the money

4.3 Events and experiences

There is a substantial body of empirical evidence to the effect that the that-clause complement of epistemic 'see' refer to *facts*, whereas the infinitival complement of perceptual 'see' refers to *events*. These are in fact the conclusions that are usually embraced in the linguistic literature (Higginbotham 1983 and the references cited there).

The first conclusion seems fully warranted. But there are some serious problems with the second conclusion. In order to conveniently formulate them, let us consider again the data in (16), repeated here as (25), for the reader's convenience.

(25) a. Yesterday I saw my baby smile, but in fact it was not my baby
      b. Yesterday I saw a baby smile, but in fact there was no baby around

A sentence like *'I saw my/a baby smile'* should express *perceptual seeing* in language. This entails that the verb 'see' in (25) should refer to an act of visual perception on the part of the relevant cognitive agent. It also entails that the infinitival complement of 'see' should refer to an event in the world (an event in which my/a baby smiles), conceived of as the object of the act of perception. Remember in fact that vision has been modeled as a cognitive event in which a distal object produces, according to the relevant environmental conditions, a proximal stimulus that sets the receptors answer in motion. A sentence such as 'I saw my baby smile' should then be analyzed as the report on a situation involving both an external event in which my baby smiled and a correct cognitive act of perception on my side, according to which I was able to turn the proximal stimulus produced by the baby smiling into a faithful visual representation of that distal object – that is, the event of smiling by my baby.

On these grounds, it follows that a sentence like 'I saw my baby smile' should not differ from a sentence like 'I saw that my baby smiled' in at least one important respect: denying that the fact referred to by the that-clause holds in the case of epistemic seeing and denying that the event referred to by the infinitival clause really took place in the case of perceptual seeing should invariably give rise to a contradiction. Namely, if it is true that the visually justified knowledge of a fact is bound to entail the very existence of that fact, it is not less true that the correct perception of a certain event is bound to entail the very existence of that event, as the distal object that is responsible for the relevant proximal stimulation.

As we have seen, this prediction is not borne out. The sentences in (25) are *not contradictory* at all. In fact, they represent the linguistic report of an optical illusion (25a) and the linguistic report of a hallucination (25b).

Notably, similar kind of reports are not possible with epistemic seeing. The reasons of this asymmetry between epistemic and perceptual seeing should be elucidated. If we want to avoid unpleasant stipulations or unnecessary complications, there is only one line of analysis available: the semantics of perceptual 'see' should not be qualified in terms of *perceptual* seeing after all. Rather, what we have so far identified as perceptual 'see' should be more correctly re-analyzed in terms of *experiential* seeing. What this means is that (25) should be analyzed as the linguistic report on a visual experience in which a baby smiled. Language is neutral about whether this visual experience reflects a correct act of perception or a case of hallucination or optical illusion. In all these cases – independently of whether they report on correct acts of visual *perception* – the *phenomenal* properties of the visual experience are exactly the same. Remember in this respect Siegel's suggestion that the content of correct acts of perception and the content of illusions/hallucinations are *phenomenally* indistinguishable.[23]

These data and observations suggest two main conclusions:

i. What is commonly referred to as perceptual 'see' in fact refers to visual *experiences*, not to acts of correct visual perception;
ii. The interpretation of the infinitival complement of 'see' does not involve existential quantification over events; the complement describes the phenomenal content of a visual experience, and that's all.

---

23 As we saw in Chapter 2, this provided an argument for Siegel's internalism.

With these two conclusions in mind, the 'logic' of the sentences in (25) becomes entirely perspicuous. For instance, (25a) states that there was a visual experience in which my baby smiled, and that this visual experience did not correspond to an event in which my baby smiled, as would have been the case with correct visual perception. For some reason, I miscategorized what I was perceiving and this resulted into a 'deviant' visual experience.

This analysis has two important consequences that should be briefly discussed. The first is that it can be elegantly modeled, linguistically, in terms of *conversational implicatures*. Normally, visual experiences whose content is φ correspond to events in which φ is the case. If I saw my baby smile, it is normally the case that my baby smiled. Suppose that this state of affairs triggers the implicature that when I report linguistically on a visual experience, what I 'say,' to make use of the Gricean terminology, is that there is a visual experience whose content is φ, but what I *mean* is that there is an event φ that brought about the visual experience. This seems entirely correct: after all, the continuations in (25) are somehow unexpected (they turn the visual experiences referred to into 'deviant' cases of visual experiences). Still, they do not make (25) contradictory. This situation is effectively captured by supposing that what (25) amounts to is just an instance, linguistically, of the common process of *implicature cancellation*. What is canceled is the implicature that having an experience whose content is φ revolves around the existence of an event φ as the distal event causing the relevant experience. On the other hand, 'implicature cancellation' is possible because it is *not* the case that experiential seeing automatically entails the existence of the content of the relevant experience as an independently existing event: the existence of that event depends on the satisfaction of further reliability conditions concerning the relation between distal objects and proximal stimulation.

The second consequence is that the proposed re-analysis of perceptual 'see' in terms of experiences (rather than in terms of perceived events) fully preserves the empirical value of the observations made above around the distinction between epistemic seeing and what we originally regarded as perceptual seeing. To just provide the explicit discussion of one case, we saw that negation inside the infinitival complement of 'see' cannot be assigned wide scope, since this move would make the complement refer to a negative event, which by definition cannot be perceived. Nothing changes,

in this respect, if an ontology of *events* is replaced with an ontology of *experiences*. Granting wide scope to the negation inside the complement of experiential seeing would be tantamount to assuming that we can have experiences without content, a not less doubtful move than proposing that we can perceive negative events.

If the analysis above is on the right track, perceptual seeing does not exist in English. What exists is *experiential* seeing, though when one employs 'see' with its experiential meaning, one normally *implicates* (although does not deductively entail) the existence of an event of perception within which what is seen is a distal object, that is, an object or an event in the external world.

From this perspective, when one says 'I saw a dog,' the existence of the dog is not what the sentence directly expresses, but is part of its 'enriched' meaning. This is why, when one utters a sentence like 'I saw a ghost' or 'I saw a unicorn,' she is not taken to suggest that ghosts and unicorns exist (or, alternatively, to talk gibberish) but promptly taken to report on some kind of hallucination or other pathological visual experience. The point is that the implicature can be canceled and 'see' can easily be taken to refer to a visual experience in which the reliability conditions for 'correct' visual perception have not been satisfied, and there is thus no distal object or event whose existence is implicated.

There is independent evidence that the object of 'see' never describes an 'external' object that is perceived. Moltmann (2013) provides an interesting discussion of some of these issues. Given sentences such as 'I saw a ghost,' there can be no doubt to the effect that

> […] the NP complements of perception verbs on the intensional reading do not describe the external object that may be perceived, but rather the way the perceived object appears (allowing for perceptual illusion) or perhaps describes a mere appearance (in the case of perceptual hallucination) (Moltmann 2013: 198).

A possible alternative, Moltmann suggests, would be to extend to the objects of perception verbs the treatment she proposes for the objects of intensional verbs such as 'need' or 'look for.' According to her treatment, these objects are *variable satisfiers*: they are the objects that instantiate the class of satisfiers in the minimal situation that satisfies the conditions expressed by the predicate. Take the sentence 'Bill needs a house.' The object in this sentence is the variable unique house that John comes to own in the

situation that satisfies the conditions expressed by the predicate, John's need of a house. If we apply this logic to 'see,' the object of 'see' should be the 'product' of an act of visual perception, something roughly interpreted as the satisfier of a correct act of perception. According to Moltmann, *special quantifiers* (quantifiers that quantify over non-referential objects, as in 'I need everything Bill needs') should quantify over products of acts of perception in sentences involving 'see' (as in, say, 'Bill saw something yellow/round'). But here there are insurmountable problems:

> One major difference between the notion of a sense datum and the notion of the product of an act of perception is that the product of an act of perception should have satisfaction conditions, such as conditions of perceptual accurateness, whereas this does not hold for sense data as they are commonly understood. Certainly, predicates of correctness are applicable to product nominalizations (a correct impression, a correct perception). However, they do not make much sense with special quantifiers as complements of perception verbs, on the relevant reading. Thus, (i) is hardly possible as a continuation of (ii):
>
> (i) In the distance, John saw a woman with blond hair
> (ii) ?? John saw something correct
>
>   (Moltmann 2013: 201)

The conclusion Moltmann draws is that though the objects of perceptual verbs appear to have intensional properties, there is no obvious intensional treatment for them. The proper semantic treatment of perception verbs remains thus an open problem.[24]

Notice that the analysis proposed here makes good prospects for the solution of this problem. According to this analysis, 'see' expresses *experiential* seeing and a sentence like 'I saw a ghost' introduces existential quantification over 'experiences.' Now, experiences are a special class of events, with peculiar metaphysical properties. For instance, they are only *privately* accessible (Fiorin and Delfitto 2020). If I see a dog, I can point to him in order for you to share my visual *perception*. But if I see a ghost, I cannot point to it in order for you to share my visual *experience*, or at least, I cannot do that successfully. Moreover, an experience and the experiencer of that experience are more strictly interwoven with each other

---

24 For the semantic and metaphysical difficulties raised by an adequate treatment of perception verbs, see also Mulligan (1999).

than it is the case for an event and the participants in that event. For instance, when I perceive something, I may easily incur into an error of misidentification ('I thought the baby I saw in this picture was me, but in fact it was my younger brother'). But when I have an experience, there can be no doubt that it's me (and not someone else) who is having that experience (#'I saw my baby smiling, but in fact it was not me who saw my baby smiling'). An experience wears the experiencer on its sleeves, so to speak. Moreover, experiences are quite less controversial, as objects of the world, than ghosts or unicorns. The question whether there 'really' are ghosts or unicorns is moot – or so we are used to think – but the question whether Bill 'really' saw a malevolent unicorn chasing him is a serious one, in the sense that there might be ways to decide whether he really had that hallucination or whether he was merely making fun of us. In fact, it might be the case that experiences provide somehow the bridge to non-existent objects like ghosts and unicorns. Namely, there is a sense in which these objects exist *as objects in some experience*. In fact, natural language seems to grant them some form of 'weak' existence, since the contrast between anaphorically resuming the object of an intensional verb such as 'look for' and anaphorically resuming the object of a hallucination is sharp enough. This is shown in (26):

(26) a. Bill is looking for a unicorn.??It is big and white
      b. Bill sees a unicorn. It is big and white

The general suggestion is thus that an ontology of experiences, besides remedying some of the empirical shortcomings of alleged perceptual seeing, as we have shown above, seems well equipped for a satisfactory account of the puzzling 'mixed' properties of the objects of perception verbs, which are both extensional (as we would conclude from the fact that anaphora is possible in (26b)) and intensional (as we would conclude from our difficulty to adapt to the idea that we inhabit a world full of ghosts and unicorns).

4.4 On experiencing, perceiving, and knowing

What is then the relationship between the verb 'see' and the perceptual concept of seeing? Roughly, vision as an act of perception entails that a cognitive agent *a* sees an object *o* or an event *e* if and only if the object *o*

or the event *e* activate the visual receptors of *a*. One sees a traffic sign or a dog cross the street if and only if there is a traffic sign in front of him and if a dog actually crossed the street. Vision as a visual experience is a different story: it simply entails that it feels like to me that there is a traffic sign in front of me or that a dog just crossed the street. A given retinotopic image is always mapped into the same representation, independently of whether this representation is veridical, that is, independently of the relationship between the proximal and the distal stimulus, that is, independently of whether this representation does or does not represent certain aspects of the world as it is.

From this perspective, the critical question to ask is thus the following: under which conditions does vision warrant knowledge? Suppose *a* sees the object *o* (a traffic sign) or the event *e* (a dog cross the street). Under which conditions does *a* come to know that there is a traffic sign and that a dog crossed the street?

The epistemology of perception provides us with the answer to this question: vision warrants knowledge if and only if all the conditions for *correct* visual perception have been satisfied (Jacob and Jeannerod 2003). As we have seen, these conditions include both the external environmental conditions that standardly regulate the relation between distal and proximal objects (like the laws of optics) and the internal formation laws that standardly regulate the computations from the sense data (the retinotopic image) to the final perceptual representation of a specific object or a specific event, through a series of well-defined stages (lower-level vision, intermediate-level vision and high-level vision). The fact that *a* feels like seeing *o* or *e* does not immediately provide *a* with the *knowledge* that *o* or *e* exist and that *a* is *perceiving* them. In order for *a* to acquire this knowledge by seeing, the experience of seeing *o* or *e* must be supplemented, crucially, with the satisfaction of the non-trivial set of conditions required for correct perception, as they are (tentatively) determined by perceptual psychology and, quite plausibly, by additional inter-modal perceptual and cognitive conditions.[25]

---

25 For instance, the veridicality of the representation must not be in contradiction with the data provided by the other senses and, more generally, with what we deem we know.

Given these rudiments of an epistemology of perception, what do we actually find in language? Epistemic seeing corresponds to the syntactic template '*a* saw that S_{finite}.' The proposition *p* expressed by the finite sentence S_{fin} refers to the knowledge that *a* has acquired by means of some visual experiences she had.[26] On the other side of the spectrum, experiential seeing corresponds to the syntactic template '*a* saw S_{inf}.' The infinitival sentence S_{inf} expresses the content of the visual experience that *a* had. If some of the conditions for correct perception are not satisfied, the *implicature* that this experience corresponds to the existence of a distal event *e* that caused the experience is canceled. The experience may then correspond to a hallucination, as when we say '*a* saw her baby smile but actually there was no baby around.' By uttering this sentence, one says that *a* had a visual experience that is typically induced by an event in which *a*'s baby smiles, but that there was no such an event in the reported case. Alternatively, the experience may correspond to an optical illusion, as when we say '*a* saw her baby smile, but actually it was not her baby.' By uttering this last sentence, one says that *a* had a visual experience that is standardly induced by an event in which *a*'s baby smiled, whereas in the reported case the relevant event has been miscategorized by *a* as involving her own baby.

The conclusion is that syntax does not contain syntactic templates that reflect the concept of perceptual seeing. These would be constructions related to a notion of seeing (let's call it 'see*') according to which the reports of hallucination and illusion above would turn out as contradictory (since a visual experience whose content is *S* would infallibly correspond to the existence of a distal event whose content is *S*). What syntax expresses, by means of the subcategorization templates of 'see' to which it gives rise, is *experiential* seeing, on the one hand, and *epistemic* seeing, on the other. And that's all.

Two questions naturally arise:

---

26 Remember that the content of these visual experiences need not correspond to the propositional content of *S*; for instance, *a* saw that Liz didn't leave by bike because *a* saw Liz's bike still at its usual place in Liz's house.

(i) Are these facts about how language reflects the epistemology of (visual) perception easily and fully appreciated in the philosophical literature?

(ii) How should these facts about how language encodes the epistemology of (visual) perception be interpreted within our broader inquiry on the relation between language and cognition?

Let us first consider question (i). Peacocke (2005) explicitly investigates the process by means of which one comes to possess the concept of perceptual experience. Here is what Peacocke writes in this regard:

> Aristotle held that it is by sight that you perceive that you see. The heart of Aristotle's idea seems to me right, provided that we understand it as follows: it is by sight that you know that you see. Suppose you see that
>
> That desk is covered with papers.
>
> This visual knowledge about the world gives you a good reason to make the self-ascriptive judgment
>
> I see that that desk is covered with papers.
>
> This is a transition you are entitled to make, from a conscious state you enjoy to a judgment. If a thinker comes to judge, by this means, that he sees that that desk is covered with papers, his judgment can thereby be knowledge (Peacocke 2005: 222)

This formulation is intended to express the *Core Rule* for the acquisition of the concept of perceptual experience, that is, the process by means of which one infers from seeing *S* that she knows that she sees *S*.

We want to take issue with this formulation. First of all, are we really entitled to make the transition from seeing *S* (*S* = that desk is covered with papers) to the judgment 'I see that *S*'? This transition is based on the knowledge that every time there is a visual experience there is some cognitive agent who is the experiencer of that experience (or 'privately accesses' that experience, if you prefer). Certainly, when there is an experience of seeing *S*, there is, by definition, a well-defined phenomenal state. If the experience consists in seeing that desk covered with papers, it certainly feels like seeing a desk covered with papers. However, this phenomenal state is not necessarily conceptualized as involving the existence of an independent cognitive/perceptual agent (an entity in the world). After all, when a baby sees a yellow flower, she probably is in a well-defined phenomenal state.

This does not entail that the baby is capable of formulating the judgment 'I see that that flower is yellow,' and this primarily depends on the fact that the baby does not possess a 'concept' for the phenomenal state she is in (the way in which we are given to ourselves is essentially *non-conceptual*; see Kripke 2011, Peacocke 2015). Moreover, even if we abstract away from this difficulty, problems remain. Suppose the transition from being in a certain phenomenal state to the knowledge that there is an entity that is the bearer of this phenomenal state is granted. After all, we all think this is what generally happens with us humans. If I have the visual experience that that desk is covered with papers, I am automatically granted the transition to the judgment according to which *'I see that desk covered with papers.'* The reason why this inference is correct is that, under an experiential concept of seeing, my access to a visual experience automatically involves the access to myself as the experiencer of that very same visual experience. This is thus *experiential* seeing. Peacocke's error consists in allowing a 'direct' transition from experiential seeing to epistemic seeing. In Peacocke's formulation above, the transition is in fact to the judgment 'I see that that desk is covered with papers,' not to the judgment 'I see that desk covered with papers.' The latter judgment is arguably sustained: it simply corresponds to describing my having a certain visual experience, whereby the transition is from having an experience to the fact that it's *me* who is having that experience. The former judgment, however, is not warranted: in order for my visual experience of a desk covered with papers to justify my *knowledge* that there is a desk covered with papers and that I am 'correctly' perceiving it, other epistemological conditions have to be satisfied, such as the compatibility requirements discussed above. After all, maybe it is not really papers what I see on the desk, or I might simply be hallucinating, something I might infer from my cognitive awareness that I frequently suffer from hallucinations.

In the end, a correct epistemology for visual perception must be based on the recognition that the transition from the visual experience of seeing *S* to the knowledge that we see that *S* is non-trivial at all: it involves delicate issues concerning the semantics of the first-person and the epistemology of perception – the additional conditions that need be satisfied in order for a given visual experience to correspond to a *correct* act of visual perception and in order for it to result into a state of justified belief.

As everyone can see, these issues are rather intricate, as it should be expected, since the epistemology of perception is a complex philosophical topic. A complex bridge has to be built between the two extremes represented by the concept of visual experience and by the concept of visually based knowledge. Even outstanding philosophers may incur in oversimplifications here.

Syntax, however, clearly picks up and unambiguously expresses these two extremes. As we have seen, a careful study of the syntactic templates that correspond to the subcategorization properties of 'see' makes it clear that these templates are designed to encode *experiential* seeing and *epistemic* seeing. No concept of perceptual seeing is syntactically encoded. If this were the case, the distinction between different ways of seeing would be somehow concealed in language, since 'correct' perceptual seeing is nothing else than epistemic seeing, after all, and this is what Peacocke's reasoning above essentially revolves around. However, experiential seeing need not be 'correct' perceptual seeing, and this is what we should not forget.

In a way, *grammar* manages to encode these subtle epistemological distinctions. And even if we had to conclude (differently from what we have concluded here) that these distinctions are in partial conflict from the conclusions we reach by means of inquiring into the epistemology of perception, the question to be asked still remains: Why should grammar care? Why should we have dedicated syntactic patterns for the expression of epistemological distinctions we are certainly not consciously aware of as non-professional philosophers and about which philosophers and cognitive scientists are often not even in agreement?

The second question we raised above is: How should the facts about the way in which language encodes the epistemology of (visual) perception be interpreted within our broader inquiry on the relation between language and cognition?

The formulation of a theory of perception and of the distinction among different ways of seeing is part of what we do within our science-forming capacity. The fact that some of the concepts we formulate at that level are reflected by grammar should fill us with a sense of bewilderment. We are not surprised to discover that the concepts evoked by words in human languages correspond to a sort of naïve metaphysics (what linguists and

philosophers often refer to as 'descriptive' metaphysics). We are also certainly not surprised to discover that the functional architecture of language expresses subtle abstract distinctions among tense, aspect and modality (Chapter 3). Conversely, to discover that the syntactic patterns associated with 'seeing' neatly reflect the high-level epistemological distinction between experiential and epistemic seeing is a different story and, for intuitively obvious reasons, somehow unexpected. It is not the same as claiming that the two types of syntactic complements of 'see' have two different interpretations (say, as propositions and as events). It really means that the interpretive properties associated with these two syntactic types neatly reflect the whole constellation of properties associated with epistemic seeing and experiential seeing. In Appendix B below we will briefly consider some of the consequences this observation has for the common semantics associated with the complements of 'see.'

As for now, let us emphasize that in chapters 1 and 3 we saw that *direct reference* is found both in visual perception and in language. In the present chapter, we have seen that syntax pre-encodes some of the subtle distinctions revealed by the epistemology of perception. These observations seem to suggest that some of the principles that are responsible for our capacity to objectively represent the world are, to a certain extent, operative across different modules of perception and cognition. This surprising finding – if correct – should not be interpreted as conflicting with T. Burge's anti-individualism. We are definitely not claiming that in order to consciously possess the concept of 'flower' we should consciously possess the higher-level concept of 'concept.' In fact, what we are tentatively arguing for is a radical version of Burge's anti-individualism. It is not only the case that one can possess and be able to use the concept of 'flower' even if one is not generally conscious of dealing with concepts. It is also the case that some of the higher-level concepts and principles that are generally deemed to be a prerogative of high cognition (say, the science-forming capacity) are in a sense already operative, unconsciously, within completely independent modules of cognition (say, language syntax). In fact, we believe that the case for what we dubbed the Harmony Thesis can be further strengthened. This is what we will see in the next chapter.

Appendix A. A note on language and non-conceptual seeing

In a classical demonstration of inattentional blindness (Simons and Chabris 1999), subjects were asked to count the passes of a basket-team during thirty seconds. In the meanwhile, an actor dressed with a gorilla suit entered the scene and beat his chest in full sight. However, many of the subjects failed to see the gorilla: inattention produces temporary blindness. Of course, it is virtually impossible that the subjects do not see the gorilla; rather, what this and many similar experiments demonstrated is that inattention prevents what they see from entering consciousness, or causes what they see to rapidly vanish from their consciousness (Dehaene 2014).

Inattentional blindness (or unconscious seeing) intuitively belongs to the same class of *non-conceptual* seeing to which we ascribed Dretske's 'simple seeing' or the cases of action-guided visual reference discussed in Chapter 3 (crucially excluding the case of Pylyshyn's FINSTs, in which non-descriptive reference is clearly associated with the conscious perception of one or more objects). What all these cases have in common is that there is no conscious experience of seeing: either no experience at all, as in the cases of inattentional blindness in which the subjects simply fail to see something that cannot have escaped their sight, or at least in part, as when we report seeing something but could not say what it is (we cannot report 'seeing' in terms of 'seeing as').

Now consider sentence (27) in English (from Gisborne 2010):

(27) Bill saw a sign but he didn't notice it at the time

For many English speakers, this sentence is not contradictory at all! But how is it possible to see something without noticing it – that is, while failing to recognize that we are having the visual experience of seeing something? And yet what a sentence such as (27) seems to do is providing a reliable verbal report of what is going on in the cases of inattentional blindness introduced above: someone sees the gorilla (how couldn't she?) but, in a way, that is disconnected from what we would define as a lasting (conscious) visual experience.

What does the non-contradictoriness of (27) tell us? At the center of the picture, we have the *perceptual cycle*: a distal object causing a proximal object (a retinotopic image) to form in the retina, and computational

processes internal to the mind that transform these patterns of distribution of the light in the retina into a conscious experiential state (a state in which, say, we can report that we 'see' the gorilla because it *feels like* seeing the gorilla).

If the perceptual cycle takes place correctly, we end up seeing something (an object, an event or a complex visual scene composed or more objects and more events) because there is something to be seen. In this case, our visual experience corresponds to the existence of the objects/events that are seen (though these objects and events undoubtedly also reflect the peculiar modes in which our mind, in a phylogenetically determined way, interacts with the physical variables of the external world). In other words, if the perceptual cycle unfolds correctly, visual experience correctly feeds knowledge, as discussed above. What we see corresponds to something that exists, and what we see can correctly feed our inference systems. This is what we called *epistemic seeing*, and syntax expresses it through dedicated formal templates (in English, that-clauses exhibiting factive properties).

Yet, something may go wrong in the perceptual cycle. We may have visual experiences without there being a distal object that brings about the process (hallucinations), we may have visual experiences in which some of the 'reliability conditions' of perceptual psychology are violated (optical illusions) or we may have distal objects correctly inducing a certain retinotopic image in the retina without being able to translate that image into the expected visual experience (as when we see a sign without noticing it, as verbally expressed by (27)).

What the verb 'see' does in English is expressing that we may *know* something by exercising correct perceptual seeing, that we may have *deviant* visual experiences (experiences that do not reflect the existence of distal objects or that do not reflect the properties of these distal objects correctly) and that we may visually *perceive* something (in the sense that light brings to our retina information concerning a certain distal object) without having the corresponding conscious visual experience.

The case in (27) illustrates the last possibility and thus corroborates the conclusions reached in the present chapter. An in-depth investigation of the syntax of 'see' in English reveals that its polysemy, as at least partially reflected by its complementation and subcategorization properties, appears to decompose the perceptual cycle in all its components: there is a 'see$_1$'

referring to visual experiences without distal objects, a 'see$_2$' referring to distal objects without visual experiences, and a 'see$_3$' referring to the knowledge we may build by performing correct acts of visual perception.

The syntax of 'see' does not simply express that we *see* what there is out there, as in a sort of folks-psychology reflection of the concept of vision. It rather works as a sort of mirror of the epistemology of visual perception, something that may engender bewilderment, as we already emphasized above.

Appendix B. Simple seeing, perceptual content, and transparency

The analysis of naked infinitival complements of 'see,' as in 'John saw Hillary kill Bill,' as instances of experiential seeing, entails that these are not standard *extensional* contexts. In the standard analysis of perceptual seeing, naked infinitivals have been regarded as referring to events. If perceptual seeing dissolves into experiential seeing, events must be reinterpreted as experiences. Experiences are a special class of events, they are events in the mind. And they give rise, as seen above, to 'intensional effects.' The re-analysis that we have proposed of perceived events as experiences entails that Higginbotham's influential extensional approach to perception verbs is mistaken (Higginbotham 1983, as a response to Barwise 1982). Higginbotham pledges for the following three properties of the naked-infinitival complements of *see*, as in 'John saw Hillary kill Bill':

  (I) Veridicality;
 (II) Referential transparency;
(III) Lack of scope ambiguity of quantifiers.

Let us begin with (I), Veridicality. Higginbotham defines it as follows:

> (I) If the unsupported clause S is quantifier-free and if S' is the present-tense full clause corresponding to S, then (i) is true: (i) If John sees S, then S'.

This cannot be right. It entails that 'John saw a ghost enter the room' warrants 'A ghost entered the room.' In the present terms, the unsupported clause S used as complement of 'see' may contain names or descriptions of objects that are *phenomenally*, and not perceptually, identified. In its semantic kernel, S counts as the verbal report of a phenomenal state. It

follows that entertaining the proposition expressed by S is to be in a different propositional attitude with respect to believing/asserting the proposition expressed by S'. The reason is that the phenomenal state expressed by S is allowed to feed belief only if that phenomenal state is judged as a correct act of perception, that is, only if a whole set of further assessment conditions in perception and cognition are satisfied.

We find that an effective way of describing what is going on in these cases is resorting to the logic of demonstratives sketched in Hintikka (1998).[27] In a nutshell, Hintikka proposes that the identity conditions for the objects referred to by demonstratives are different from the identity conditions that hold for names and descriptions. There are thus two distinct domains of quantification, one for names and descriptions (this is the domain of 'public' objects), and one for demonstratives (this is the domain of 'perspectival' objects). Hintikka's central insight is that when I point to a building while uttering 'This is the Empire State Building,' the object referred to by the demonstrative is something that is given in perception but whose identity in the publicly accessible, non-perspectival sphere is still not established. There are in fact possible worlds in which the perspectival object referred to by 'this' is not the Empire State Building but, say, the Trump Building. In a sense, I am thus quantifying over entities that are objectively given in perception (when I utter the sentence, my interlocutor is plausibly seeing the same building I am seeing) but not objectively given in cognition yet, since a perspectival object (that is, a perceived object) may be mapped into distinct public objects in different worlds.

If we apply Hintikka's logic to the example 'I saw a ghost enter the room,' the idea is then that we are quantifying over objects of perception, along the lines of (28), which should be accurately kept apart from (29).

(28) $(Ex) K_I ((ghost(x) \& entered\text{-}the\text{-}room(x))$
(29) $(\exists x) (x=b) \& (Ey) K_I ((ghost(y) \& entered\text{-}the\text{-}room(y) \& (x=y))$

What (28) expresses is that there is an object of perception x (a perspectival object in Hintikka's terminology), such that, in my act of perception, x is a ghost and x entered the room. In the case at stake, we know that there

---

27 Hintikka's approach is essentially incompatible with Kaplan's, though we will not discuss these issues in any detail here.

is no 'public' object corresponding to this perspectival x. If perspectival y could be mapped into a 'real' object x, whose identity could be publicly ascertained and accepted, then what we would get is (29), according to which the object of perception that looked as a ghost and entered the room is in fact a public object *b*. In other words, (29) is the logical form expressing the assertion that the ghost that was *phenomenally* given to me as entering the room really exists and really entered the room. My phenomenal experience was a publicly ascertainable event.

In general, as proposed above, naked infinitivals as complements of 'see' (Higginbotham's unsupported S) trigger the implicature that the domain of quantification is that of real objects. This is the reason why *If John sees S, then S'* is generally true. However, as in the case just discussed, this implicature can be cancelled. In this case, there is no mapping from the domain of perspectival objects to the domain of real objects.

Now, what happens exactly when the implicature is not cancelled, as when we verbally report a perceptual act (and not a hallucination) by uttering 'John saw Hillary kill Bill'? In this scenario, the event of killing is not purely phenomenal, and both Hillary and Bill are public entities, real participants in a real event, so to speak. In this case, when we verbally report that 'John saw Hillary kill Bill,' we are plausible ascribing to John a propositional attitude of belief towards S. Entertaining the perceptually given content *p* expressed by S in a context of correct perception is thus tantamount to being in the propositional attitude of believing *p*.

Experiences correlating to correctly perceived external events have a different propositional content than purely phenomenal experiences. The entities (objects and events) involved in the latter are objects of perception (perspectival objects), whereas the entities (objects and events) involved in the former are public objects, whose identity is established in a shared cognitive space. From this semantic perspective, we do not agree with Siegel. The propositional content of hallucinations is distinct from the propositional content of correct acts of perception: it is only in the latter that perspectival objects are mapped into public entities (objects and events), objectively supporting states of belief.

However, this is still not the whole story. A belief-supporting perceptual report – say, 'John saw Hillary kill Bill' – remains inherently ambiguous between a perceptual experience of 'seeing as' and a perceptual experience

of 'simple seeing,' in Dretske's sense. When I utter 'John saw Hillary kill Bill,' I intend that John had a perceptual experience compatible with a belief-supporting propositional representation. I intend that what John saw is compatible with the propositional representation 'Hillary killed Bill,' and could legitimately feed John's belief that Hillary killed John. The ambiguity consists in the fact that in uttering the sentence I am not necessarily asserting that what John saw *actually* fed his belief that Hillary killed Bill. Suppose I say 'John saw Hillary kill Bill, but John did not realize that she was killing him.' This makes perfectly sense. What I mean by that is that John visually witnessed a killing, without ever properly categorizing it as a killing, that is, in Dretske's sense, without seeing it *as a* killing. In other words, what I am reporting is an instance of 'simple seeing': one sees the gorilla without being aware of seeing it; analogously, John sees a killing without being aware of seeing it.

Still, when I verbally report that 'John saw Hillary kill Bill,' I do not report a hallucination or an optical illusion. What I report is a visual experience grounded in an act of correct perception. The fact that John's seeing *p* may not feed his belief that *p* is the case simply depends on the semantics associated with *de re* interpretations. Consider in this respect Quine's 'double vision.' John sees a man wearing a brown hat in the street and get persuaded that this man is a spy. In a different moment, John meets a grey-haired man he is acquainted with as a pillar of the community and of whom he thinks – of course – that he is not a spy. In fact, unbeknownst to John, the two men are the very same person, a guy called Ortcutt. This complex situation can be reported by means of the two sentences in (30) and (31):

(30) John believes that Ortcutt is a spy
(31) John believes that Ortcutt is not a spy

It follows that John should believe a contradiction, that is, he should believe that the very same person (namely Ortcutt) is a spy and isn't a spy. Plainly, this is not the case. The puzzle is solved when we consider that a *de re* reading remains neutral about the way in which the objects referred to (in this case, Ortcutt) are perceptually/cognitively given to John. In (30) Ortcutt manifests to John as the man with a brown hat, whereas in (31) Ortcutt is given to John as the good grey-haired man everybody is acquainted with in

the community. The modes of presentation involved are utterly different. As a consequence, John need not be aware that these two ways of perceiving an individual are two distinct ways of perceiving the same individual.

On these grounds, when we verbally report an act of perception that might legitimately feed John's belief that Hillary killed Bill, we are not committed to imply that what we *see as* an event of killing – a public object in Hintikka's sense – is something that also John sees as an act of killing. In fact, it is perfectly possible that John is not aware that what is going on in the scene he visually witnesses is a killing, exactly as John may not be aware that what he is visually witnessing is a gorilla amidst a team of basket players. In other words, and quite interestingly, we use names and predicates referring to public objects in order to describe what is in fact an instance of Dretske's 'simple seeing.' From this perspective, though, simple seeing emerges as the result of a form of correct, though perhaps incomplete, act of perception, whereby the incompleteness is a consequence of the lack of mapping between privately perceived objects (perspectival objects in Hintikka's terminology) and real objects whose identity conditions are publicly established and shared.

As a conclusion, it seems clear that the issues about Higginbotham's veridicality are many and complex, and the story to tell, once we do justice to the distinct ways of seeing, is more intriguing and fascinating than it may appear *prima facie*. At the end of the story, anyway, one should acknowledge that veridicality does not hold and that objects of seeing are not extensional, at least not according to the standard interpretation of extensionality.

Let us now consider (II), Referential transparency. Higginbotham defines it as follows:

> (II) The context V ___ is referentially transparent, if V is a perception verb and if what fills the blank is an unsupported clause (that is, a naked infinitival).

As an empirical observation, (II) is correct. It entails that 'John saw Hillary kill Bill' warrants 'John saw the first-lady kill the president.' However, this referential transparency effect cannot be used to argue for an extensional analysis of the infinitival complements of perception verbs. In fact, the reason why (II) holds are not different from the reasons why one can truth-faithfully utter 'John believes that the first-lady killed the president'

even in a context in which she knows that John is not aware of the fact that Hillary is the first lady and Bill the president. The reason is that the relevant descriptions can be read *de re*. Under a *de re* reading, the mode of presentation 'the first lady' is the way the relevant individual (Hillary) is publicly identified by the speaker, whereas it is not necessarily the way in which John identifies it.

The same situation arguably holds for perception verbs. I can truth-faithfully utter 'John saw the first-lady kill the president' even in a context in which I know for sure that John knew nothing about the institutional roles of Hillary and Bill. In this case as well, the relevant descriptions are read *de re*, to the effect that the relevant mode of presentation is ascribed to the speaker and not to John. We conclude thus that the alleged referential transparency of perception verb complements is in fact a by-effect of the *de re* readings of the descriptions contained in these infinitival complements. There is no reason to pledge for their extensional status, unless we are prepared to concede that the substitution of 'Hillary' with 'the first lady' under a *de re* reading of 'John believes that the first lady killed Bill' provides an argument for the extensional status of *believe*-contexts.

Finally, let us consider (III), lack of scope ambiguity of quantifiers. According to Higginbotham, this boils down to the following: All conditionals of the sort of (IV) are true: (IV) If John sees somebody leave, then there is somebody whom John sees leave

Needless to say, we do not accept the truth of (IV). If (IV) held, it would entail that the infinitival complements of perception verbs are always quantified in from outside, which is definitely not the case. Technically, in the present terms, the function from the 'perspectival' objects of perception to the public objects of general cognition is a *partial* function.[28] Empirically, (III) simply obliterates the obvious fact that incorrect perception can be not only illusions but also hallucinations.

---

28 A partial function is a binary relation defined over two sets such that every element of the first set is associated with *at most* one element of the second set. In the case at stake, there is not necessarily a public object for each perspectival object.

Chapter 5   Does syntax tell us what propositions are?

5.1 Silent nouns and relative clauses

In recent work in comparative syntax, Kayne has proposed that a large variety of syntactic constructions contain a number of 'silent' heads. These are constructions in which the head of certain phrases is left unpronounced and can thus be detected only by means of indirect empirical evidence, involving the application of sophisticated heuristics. Kayne produces original and intriguing arguments in favor of the view that at least the following constructions in English and French contain a silent head, which is given here, following Kayne's orthographical conventions, as the noun in capital letters (Kayne 2010):

(1) a. John has fewer books than Bill → John has fewer NUMBER books than Bill

b. John has bought a few houses this year → John has bought a few NUMBER houses this year

c. Jean   achetait   des   livres → Jean achetait NUMBER de les livres
Jean   bought   of-the   books

d. Jean   buvait   de   la   bière → Jean buvait AMOUNT de la bière
Jean   drank   of   the   beer

e. John went there → John went there PLACE

f. Jean   y   va → Jean y PLACE va
Jean   there   goes

g. John spoke thereof → John spoke there THING of

h. Jean   y   pense → Jean y THING pense
Jean   there   thinks

i. A red COLOR car (that car is red in color)

l. Therefore → There REASON fore

As the reader may have noticed, all these silent heads correspond to abstract nouns, that is, abstract concepts endowed with a non-trivial cognitive import (amount, number, thing, color, reason, place). One of the cases discussed by Kayne has potentially far-reaching consequences for the

hypothesis, considered in Chapter 4, according to which syntax pre-encodes some subtle epistemological distinctions that explicitly emerge only through conscious rational inquiry. In what follows, we intend to present and discuss this case in some detail.

The starting point is Kayne's analysis of constructions such as (2), which are canonically interpreted as involving a sentential complement (a that-clause) selected by the noun 'fact.'

(2) the fact that p

Kayne considers some reasons why the head-raising analysis he has proposed for relative clauses might be appropriate for (2) and in fact preferable to the traditional idea that 'that' in (2) is a 'complementizer' that introduces a clausal complement. Let us first briefly see what the head-raising analysis of relative clauses amounts to and then briefly review what reasons there are to analyze 'that p' in (2) as a relative clause.

Consider the case where a noun selects a relative clause, such as in (3a). According to Kayne's analysis, a DP such as 'the fact that you noticed' should be assigned the representation in (3b), where a base-generated determiner-head selects a CP:

(3) a. the fact that you noticed
     b. [$_{DP}$ the [$_{CP}$ you noticed [$_{NP}$ that fact]]]

The word-order in (3a) is derived from the syntactic representation in (3b) by means of two movement-steps. The first involves displacing the NP 'that fact' to the left-periphery in (3b), the second involves displacing the noun-head 'fact' to the edge of CP. According to the labeling convention adopted in Cecchetto and Donati (2015), the displaced noun projects as a NP, giving rise to the hierarchical representation in (3c), where the base-generated determiner 'the,' as the result of the syntactic derivation, correctly selects a NP instead of a CP (as in (3b)):

(3) c. [$_{DP}$ the [$_{NP}$ fact [that fact [$_{CP}$ you noticed that fact]]]]

With this in mind, suppose now that (2) corresponds to the full noun phrase in (4a):

(4) a. The fact that Liz left

What are the reasons to propose that the complement clause 'that Liz left' is a relative clause? If this were the case, (4a) should be analyzed as involving something like the representation in (4b), a *prima facie* bizarre analysis:

(4) b. [$_{DP}$ the [$_{NP}$ fact [that fact [$_{CP}$ Liz left that fact]]]]

However, consider the sentence in (5a). Uncontroversially, this is a relative clause that is associated, under a head-raising analysis, with the structure in (5b):

(5) a. The way in which Liz left
     b. [$_{DP}$ the [$_{NP}$ way [in which way [$_{CP}$ Liz left in which way]]]]

Now, it is just a fact of English that (5a) can also be expressed as (5c):

(5) c. The way that Liz left

According to Kayne, this observation suggests that the relative pronoun 'that,' in English, can simply stand for the complex relative pronoun 'in which.' If this is the case, it suggests in turn that (4a) might correspond to (4c), whereby 'that' has been replaced by 'in which':

(4) c. The fact <u>in which</u> Liz left

The conclusion would be that 'that' is a relative pronoun in (4a). What appears as a sentential complement might actually be a relative clause. This might be the case if (4a) involves the presence of an empty preposition IN. And this may not be so difficult to accept, if only we consider the singular correspondence between an expression like 'the fact that Liz left' and an expression like 'Liz left, *in* fact,' where the implicit 'in' is no longer silent. Under the relative clause analysis, the connection between the two structures becomes transparent: both structures involve the preposition 'in,' though it is silent in one of them.

If correct, Kayne's arguments that a structure such as (2) involves a concealed relative clause raise a general issue regarding the nature of the sentential complements of some specific classes of verbs, including *factives* (6a), *verbs of perception* (6b), and *verbs of propositional attitude* (6c).

(6) a. John regrets that Liz left
     b. John saw that Liz left
     c. John believes/thinks that Liz left

It is clear that if one could assume that the structures in (6) contain an empty noun such as FACT, it would become rather natural to analyze the alleged sentential complements in those structures as involving a concealed relative clause. As a consequence, 'that' in (6) would no longer count as a 'complementizer' introducing a clausal complement. As in (2) above, 'that' should be analyzed as a relative pronoun.

Interestingly enough, in spite of its *prima facie* counterintuitive nature, this hypothesis would immediately resolve a long-standing issue in the philosophical discussion of propositional attitudes, where it was explicitly asked what the nature is of the 'complementizer' that introduces the clausal complements. The homophony detected in English and in other languages between this alleged complementizer and one of the demonstrative pronouns ('that') led to the so-called *paratactic theory*, according to which a sentence like (7a) should be analyzed along the lines of (7b). This analysis, originally championed by D. Davidson, has also been adopted by linguists who argued for a diachronic analysis in which complementizer 'that' was originally a demonstrative pronoun (Roberts and Roussou 2003):

(7) a. John believes that Liz left
    b. John believes that. Liz left

The paratactic theory induces an unresolved puzzle. Though (7a) is naturally paired by (7b), (7c) cannot be associated with the entirely unacceptable (7d):

(7) c. John believes this. Liz left
    d. *John believes this Liz left

The fact that *this/that*, though naturally paired as demonstratives, are not paired as complementizers, is further manifested by the paradigm in (8) (Kayne 2010: 191):

(8) a. I'm sure that/*this you're right.
    b. That/*This you're smart is obvious.
    c. It bothers us that/*this you're right.
    d. We'll see to it that/*this he does the job

It goes without saying that if complementizers are in fact relative pronouns, the paratactic theory, as well as the concomitant association of complementizers with demonstratives, would fade away as an illusion.

5.2 Facts and events as silent nouns

In this section, we will consider the hypothesis that all types of that-clauses in (6) are in fact concealed relative clauses, together with some of its consequences. The first question to address is thus whether there is empirical evidence for the relative clause analysis, by weighing this evidence for each class of verbs. In the case of factive verbs, as exemplified in (6a), the evidence is not far-fetched. As is well-known (see also the discussion in Chapter Three), the complements of factive verbs are traditionally associated with a presuppositional analysis in which the truth of the proposition that they express is a pre-condition for the sentence as a whole to be assigned a truth-value. If 'John regrets that Liz left' is true, then that Liz left is also true. And if 'John regrets that Liz left' is false, this also invariably entails that Liz left. Given (6a), that Liz left must be a fact. Here is how Kayne interprets factivity in syntactic terms:

> An idea that goes back to Kiparsky and Kiparsky (1970) is that factive sentences like:
>
> (i) We're sorry that you're here.
>
> have a deleted or silent FACT. If so, […] then factives, too, must involve relative clause structures (based on 'IN FACT'). This provides an immediate account of:
>
> (ii) *We're sorry this you're here.
>
> in terms of the inability of to serve as a relative pronoun […]. That factives are relative clause structures is supported by the observation in Roberts and Roussou (2003: 120) to the effect that in Modern Greek the relative clause marker 'pou' is also used with factives.
>
> (Kayne 2010: 216).

In other words, the really difficult step is to show that structures of the kind in (2) are prone to an analysis as relative clauses. Once this is established, the extension of this analysis to the complements of factive verbs is essentially trivial.

More particularly, a relative clause analysis crucially involves head-raising, and head-raising is only possible if there is a head-noun to raise. This means that sentences such as (6a) should be re-analyzed along the lines of (9a) below, whereby FACT is a silent noun of the sort of those briefly reviewed in (1) above, and which gives rise to the syntactic structure in (9b):

(9) a. John regrets the FACT in which Liz left
    b. John regrets [$_{DP}$ the [$_{NP}$ FACT [in which FACT [$_{CP}$ Liz left in which FACT]]]]

Factive verbs are thus relatively easy to deal with. What about the sentential complements of *perception* verbs? Though the factive nature of these complements is often left undecided in the linguistic literature, in Chapter 4 we have provided substantial empirical reasons why these complements should also be interpreted as involving an implicit reference to *facts*. The crucial factor to be considered is the *epistemic* interpretation of these that-clauses. What they express is not the content of a visual experience (as in 'John saw Liz leave') but the *knowledge* that the cognitive agent has of a fact (here, the fact that Liz left). The epistemological divide between facts, on the one hand, and events/experiences, on the other, has been argued to be of crucial importance to explain the inferential behavior of the speakers of English when they produce/interpret the different syntactic templates associated with perception verbs. To put it shortly, the syntactic analysis of sentence (6b) along the lines of sentence (10a), whose full structure is rendered in (10b), is supported by the semantics associated with the presence of the silent head FACT in (10a-b), based on the insight that perception verbs are actually factives:

(10) a. John saw the FACT in which Liz left
    b. John saw [$_{DP}$ (the) [$_{NP}$ FACT [in which FACT [$_{CP}$ Liz left in which FACT]]]]

We conclude that the application of a relative-clause analysis to the alleged sentential complements of (6a) and (6b) is rather trivial, at least once we have adopted this analysis for the structure in (2). This approach is tantamount to extending Kayne's insights about the pervasiveness of silent heads in language to all clausal complements of *factive* predicates, crucially including perception verbs such as 'see.' Significantly, a relative-clause analysis goes hand in hand with the presence in syntax of unpronounced nouns that correspond to abstract concepts with a specific ontological import.

At this point, the obvious question to be raised is whether a relative-clause analysis, involving the presence of a silent head, can be extended to the alleged sentential complements of verbs of propositional attitudes, exemplified in (6c).

Here, some serious problems arise. It makes no sense to propose that (6c) contains a silent head FACT, as it is arguably the case with (6a) and (6b). 'Believe' is not a factive predicate. Sentence (7a), for instance, does not convey the presupposition that Liz left; in fact, whether (7a) is true is completely independent of whether Liz left. This is confirmed by the observation that it is verbs of propositional attitude such us 'think' and 'believe' that standardly give rise to *de re/de dicto* ambiguities, as we saw in Chapter 4. Suppose Liz is Bill's daughter; the propositional content proper to (7a) can then be expressed as in (11):

(11) John believes that Bill's daughter left

This sentence can be interpreted *de re*: there is an individual who is Bill's daughter, namely Liz, and John believes of that individual (Liz) that she left. However, (11) can also be interpreted *de dicto*: John believes that a person being Bill's daughter left, without knowing that this person is in fact Liz. As we have seen, factive predicates are essentially immune to *de re/ de dicto* ambiguities. On these grounds, to propose that the constructions exemplified in (6c) contain FACT as a silent head would be completely unwarranted.

What should we conclude? If there is no independent evidence for the presence of a silent noun in the complements of 'believe,' this should be taken to entail that a relative-clause analysis of these complements is untenable. On the other hand, it would be bizarre to propose that while 'regret' and 'see' select for a relative clause (and 'that' is a relative pronoun with these verbs), the that-clause selected by 'believe' is a sentential complement (and 'that' is a complementizer in this case). We would no longer understand, for instance, why 'believe' does not select both 'that' and 'this' as complementizers. Whatever they are (sentential complements or relative clauses), all these that-clauses should receive a unique syntactic analysis, as a rather obvious requirement for explanatory adequacy.

Kayne is fully aware of these consequences, and proposes a solution aimed at rescuing a relative-clause interpretation of the that-clause in (6c). His suggestion is based on Rosenbaum's old conjecture that 'believe' takes a silent pronominal IT as its object ('John believes IT that Liz left'). If this is the case, IT would be the raising nominal and a relative-clause interpretation of the complement of 'believe' would be feasible, after all.

However, we think that there are serious problems with this proposal. First of all, whereas an overtly realized 'it' does not change the factive interpretation of 'regret' (we have both 'John regrets that Liz left' and 'John regrets *it* that Bill left'), the presence of an overtly realized 'it' with 'believe' yields a sensible change in interpretation. This becomes particularly clear thanks to minimal pairs as in (12):

(12) a. I didn't believe that Liz left. In fact, she didn't
     b. I didn't believe *it* that Liz left. *In fact, she didn't

The presence of 'it' in the first sentence of (12b) makes the continuation in (12b) hardly acceptable. This datum suggests that the presence of 'it' as the object of 'believe' tends to turn 'believe' into a factive verb. More precisely, it seems that the presence of 'it' produces giveness/familiarity effects, both with factive and non-factive predicates. For instance, 'John regrets it that Liz left' is natural only in contexts where the propositional content of the that-clause (the fact that Liz left) has already been introduced and is therefore familiar to the discourse participants. This also applies to 'believe': a sentence such as 'John believes it that Liz left' also presupposes that Liz's departure is something familiar in the context of utterance. All in all, what overt 'it' does in the case of 'believe' is to change the interpretation of its propositional complement (from non-familiar to familiar and, in certain contexts, as for instance (12b), from non-factive to factive). This introduces an ugly asymmetry with alleged silent IT, which expresses neither familiarity nor factivity. Silent IT must always be present when 'believe' takes a that-clause as its complement, if we want a relative-clause analysis of those complements, but the properties of silent IT are completely opposed to the properties of overt 'it.' This is in a striking contrast with the elegant analysis that we have for 'regret,' since there is no silent IT with 'regret.' The latter is always factive because of the presence of silent FACT, and the propositional content becomes 'familiar' when FACT is overtly pronominalized as 'it.'

So, which alternatives do we have to the presence of IT? Well, we may maintain that 'believe' is not factive for the simple reason that it does not select FACT, but EVENT. The syntactic structure of 'regret' and 'believe' would then be the same, but the silent head involved would be quite different in its conceptual/referential content. If 'believe' selects EVENT, a sentence such as (7a) corresponds to the more abstract (13a), to be associated

with the syntactic configuration (13b) after head-raising of silent EVENT has taken place:

(13) a. John believes the EVENT in which Liz left
       b. John believes [$_{DP}$ (the) [$_{NP}$ EVENT [in which EVENT [$_{CP}$ Liz left in which EVENT]]]]

The syntactic representation in (13b), involving silent EVENT, should be compared with the representation in (10b), involving silent FACT. What (13b) and (10b) have in common is that the raising noun is a silent head and that they both involve the silent preposition IN. It is the presence of silent IN that explains why the relative pronoun 'which' is pronounced 'that' in (13b) and (10b), based on the free alternation between 'the way in which p' and 'the way that p.' The two representations differ from each other in the nature of the silent head that is selected (FACT vs. EVENT). The next question is thus the following: Which reasons do we have to propose that 'believe' selects for EVENT?

Here, there is an unexpected association to be drawn. In Chapter 1, we have seen that Scott Soames proposes to interpret propositions as cognitive 'events' of property-ascription (in perception or in cognition). When I see a certain flower as yellow, I entertain the proposition that that flower is yellow. This proposition is the cognitive event by means of which the relevant property (yellow) is assigned to the relevant object (the flower). The same is the case when a speechless infant sees a flower as yellow – or so we have argued. If propositions are cognitive events of property-ascription, and the complement of verbs of propositional attitude such as 'believe' is a proposition, what we should expect is that 'believe' selects for an abstract concept of cognitive event. Translated in linguistic terms under Kayne's analysis, according to which the complement of 'believe' is syntactically a relative clause, and requires as such a silent noun as the raising head, this entails that we should expect that 'believe' selects for a silent noun of the type of EVENT. This neatly supports our proposal. But what does it exactly mean?

The obvious thing to say is that there are different theories of propositions (propositions as structured objects, as sets of possible worlds and as cognitive events of predication). As we discussed in Chapter 3, there are ways to show that these theories are not mutually incompatible. However, if we

accept Soames' arguments – and we find them quite convincing -, the theory of propositions as cognitive events of predication is somehow *foundational* with respect to the other notions of propositions: it is the only theory that explains (in a non-circular way) how propositions can be truth-bearers, since the truth of a cognitive event of predication is linked to the satisfaction to the relevant set of *correctness* conditions in perception or cognition, as these conditions are formulated in perceptual psychology and in the theories of cognition.

What we found is evidence that Soames' analysis is mirrored by syntax. We have seen that there are syntactic reasons to propose that that-clauses, as *propositional* complements of 'believe,' are *relative clauses*. If they are, they involve a silent head. This silent head cannot be FACT. However, it might well be EVENT, if propositions are the types of cognitive events of predication that Soames proposes. This proposal is fully compatible both with the syntax (a relative-clause analysis) and the semantics (non-factive properties, presence of *de re/de dicto* ambiguities) that we have established for the clausal complements of 'believe.'

Surprisingly enough, there are extra sources of empirical evidence in favor of the hypothesis that the clausal complement of 'believe' contains the empty noun EVENT and the empty preposition IN. First of all, the preposition 'in' overtly surfaces in one of the two non-clausal subcategorization frames associated with 'believe,' as in 'John believes *in* Bill' or in 'John believes *in* power.' Even more significant, perhaps, is the fact that the interpretation of 'believe' in the structures featuring overt 'in' arguably parallels the interpretation of 'believe' as a verb of propositional attitude, where 'in' is silent and is hypothesized to introduce the silent noun EVENT.

In order to see how this insight may be implemented, it suffices to suppose that BELIEVE is BELIEVE IN (this will be our working hypothesis). On these grounds, let us further suppose that

(14) a. x BELIEVES IN y

is semantically equivalent to

(14) b. x considers [y trustworthy].

By hypothesis (see discussion above), we know that

(14) c. x believes that p

is interpreted as

(14) d. believes in the cognitive event in which P(a)

and eventually as

(14) e. x considers [the cognitive event in which P(a) trustworthy]
          (by the equivalence between (14a) and (14b)).

It also follows that:

(14) f. x believes in John/money/art

is interpreted as in

(14) g. x considers [John/money/art trustworthy]
          (by the equivalence between (14a) and (14b), and under a context-
          dependent interpretation of the predicate 'trustworthy').

This argument supports the hypothesis that the meaning of 'believe' is exactly the meaning that 'believe' expresses when it overtly selects 'in,' witness 'John believes in Liz.' Here, 'believe' means 'to consider Liz trustworthy,' which exactly parallels the proposed interpretation of 'John believes that Liz left' in terms of 'John considers the cognitive event in which Liz left trustworthy.'[29]

On these grounds, we also come to understand why the silent head in (13a) must be EVENT and not, say, something like PROPOSITION, whatever PROPOSITION is intended to mean. In 'John believes Liz,' the meaning is not that John trusts what Liz *does*, it is rather that John trusts what Liz *says*. What John considers trustworthy is some (implicit) cognitive event of predication whose agent is Liz. The link between Liz, as the object of 'believe,' and this act of predication is relatively straightforward: Liz is the cognitive agent of the event of predication. In fact, there is no event of predication without a cognitive agent. Conversely, propositions are abstract mathematical structures, whose relationship with the cognitive agent that may entertain them is irrelevant to them, hence more indirect. If 'believe' expressed *trusting a proposition*, rather than *trusting a cognitive event of predication*, it would be troublesome to understand why English makes

---

29 We are indebted to Eric Reuland for important suggestions and feedback on
    this point.

use of syntactic structures such as 'John believes someone' to express that John trusts what someone said. Propositions, contrary to cognitive events of predication, are not defined by the presence of the cognitive agent who may entertain them. And syntax seems to reflect this fact: in 'John believes Liz,' the syntactic object of 'believe' is the cognitive agent of the implicit cognitive event of predication that John considers trustworthy.

5.3 Harmony and some of its effects

Suppose this is, essentially, the story to tell. Propositions, as expressed in language, are not abstract mathematical structures but cognitive events of predication. Syntax mirrors this format rather closely.

There is a sense – we submit – according to which this story is disconcerting. Certainly, virtually all linguists would endorse the view that morphosyntax encodes important aspects of interpretation. Tense phenomena are the grammatical encoding of the concept of 'time,' after all. 'Aspect' grammatically encodes possible ways of looking at the internal structure of events, and so on. There is nothing particularly disconcerting about the existence of dedicated morphemes, realized within well-defined hierarchical structures, which encode the notions of 'future' or 'past' or the notion that a certain event is 'ongoing' rather than already 'accomplished.' So – one might argue – there is also nothing really extraordinary in the finding that certain syntactic constituents express 'propositions' in the way modern cognitive science have identified them. After all, at the interface between syntax and the systems of interpretation, propositions are needed 'as referents of that-clauses, arguments of attitude verbs, referents of some names and uses of indexicals, members of the domains of some quantifiers, and so on' (Soames 2015: 93).

We fully agree with these observations. The sense of disconcert regards something else. It concerns the fact that syntax apparently does something more than just encode 'propositions' by means of quantifiers (when they range over propositions) and by means of names (when they refer to propositions). This would simply entail that syntax provides us with a tool for referring to propositions and for thinking about propositions, *whatever propositions are supposed to be*. But now it seems that syntax 'pretends' to tell us *what propositions exactly are*. And this is the troubling point. Let us try to elucidate it in some detail.

In Chapter 4, we saw that our use of sentences in natural language presupposes the existence of abstract objects such as *facts* and *events*. The task of *descriptive natural language metaphysics* is in fact to discover which are the perhaps bizarre objects whose existence we arguably presuppose when we speak or comprehend language. This task involves a deep and systematic analysis of language. In Moltmann's words, 'natural language ontology obviously is part of naïve metaphysics. But of course, it is not 'naïve' in the sense of concerning itself with what the ordinary person naively takes there to be. Rather, as part of the metaphysics of appearance, it deals with the ontological categories, notions, and structures that a deep and systematic analysis of language uncovers' (Moltmann 2016: 9). So, for example, by investigating the inferential properties that are manifested in language (the fact, say, that 'John solved the problem quickly' entails 'John solved the problem') we may propose that we make unconsciously use of an ontology of events, in which adverbs like 'quickly' are predicates of events, to the effect that the inference above can be modeled on the Boolean inference from 'there is a yellow flower' to 'there is a flower.' The naïve metaphysics underlying natural language involves facts and events and it is quite possible that syntax 'unconsciously' manipulates these abstract objects. If syntactic complements of 'believe' can be shown to involve, on the independent syntactic grounds that we have explored above, a silent EVENT as the nominal raising head, syntax would provide a direct and unexpected confirmation of the hypothesis that natural language ontology contains events, as well as facts. Syntax would then offer an entirely new kind of evidence concerning the descriptive/naïve metaphysics that underlies natural language.

We can now see what the nature of the issue we have raised really is: *Is this still naïve metaphysics*? When we introduced the discussion, in Chapter 1, about the choice between the notion of proposition as a set-theoretic object and as a cognitive object, this was intended as a discussion in *foundational* metaphysics. In effect, we inquired into the properties that the notion of *proposition* must have in order for it to count as the conceptual foundation of the *representational* properties of sentences, crucially including *veridicality* and the formulation of truth-conditions. We tentatively concluded that an analysis of propositions as cognitive events, within foundational metaphysics, is what appears to offer the best prospects for an adequate

account of propositions as truth-bearers, that is, as the metaphysical foundation of veridicality and truth-conditions. Discovering, on completely independent syntactic grounds, that certain that-clauses contain a silent noun EVENT that directly supports Soames' analysis of propositions as cognitive events leads us to the very same disconcerting conclusion that we have reached in Chapter 4 concerning the semantics of verbs of visual perception: it looks like syntax somehow *mirrors*, or *pre-encodes*, some of the results of foundational metaphysics. In the case of perception verbs, it was the clear-cut divide between *epistemic* seeing and *experiential* seeing. In the case of the syntax of the clausal complements of 'believe,' it is the evidence for an analysis of propositions as involving reference to cognitive events.

According to the theoretical proposal we have endorsed in Chapter 1, propositions are cognitive events of predication. Syntax directly confirms this: a sentence like 'John believes that Liz left' corresponds to the more abstract syntactic structure 'John believes the event in which Liz left,' which features two silent heads (the noun EVENT and the preposition IN). The evidence that we have discussed in this chapter supports this *prima facie* fancy idea. It is then not easy to resist the temptation to conclude that syntax, however preposterous or outrageous this may sound, has some lessons to tell in *foundational metaphysics*. If this is correct, it yields bewildering. Once again, we are faced with odd mirror effects across different cognitive modules. Just an optical illusion or Harmony Effects across the operations of the mind?

Epilogue   Language as a looking-glass and grammar as a rabbit-hole

The Cartesian tradition in the study of language has many merits. Undoubtedly, syntax is a sophisticated computational module whose defining properties (discrete infinity, recursion, locality) are entirely original or at least originally implemented within language. Unfortunately, Cartesian linguistics is also exposed to tangible risks of being confused with what T. Burge calls individual representationalism. Concepts are nothing else than lexemes, propositions nothing else than sentences, in a word, the capacity for thought is nothing else than the capacity for language. Since language is the essential vehicle for the development of all higher-level cognitive systems, language is detached, by definition, from the systems of lower cognition and, more particularly, from perception.

It is here that this flawed Cartesian view of *language* meets a flawed Cartesian view of *perception*, a view according to which sensory systems are nothing more than transduction systems and sensory consciousness nothing more than the passive registration of energy patterns that are incident upon the outer sensory receptors. Sensation is deprived of any representational power and lacks semantic significance. The true signatures of the human mind, classification and conceptualization, are the result of epistemic faculties that represent a kind of mental activity that is orthogonal to sensation. The categorizing and conceptualizing activity of the soul applies to the raw data of sensation as light impinges on physical matter. Beliefs and knowledge arise only at the moment the dark and shapeless world of sensation is turned into the representation of what is true or false, correct or incorrect, as the result of the power of thought and reason, as reflected in language.

This is an incorrect representation of how things are, which turns reality upside down. Sensation already involves complex computation. Neurons already perform classificatory activities of a *propositional* nature, by grouping sets of physical conditions into psychologically relevant equivalence classes. As neurons' response to patterns of stimulation can be modeled propositionally, visual processing (and perception more generally) can be functionally represented in terms of unconscious propositional attitudes,

and transformations among percepts in terms of unconscious inferential patterns.

In vision, different feature maps are created and then brought together to create the representation of three-dimensional space. Vision is combinatorial/compositional: the processes involved in the visual representation of a red circle and a blue square consist in binding, at a given location, one of the ingredients of a color map with one of the ingredients of a shape map, and by differently combining the ingredients of these maps we easily get, in perception or in imagination, the different representations of a blue circle and of a red square. Moreover, vision is referential: both epistemic vision and motion-guiding vision have all of the ingredients of the two varieties of reference that have been detected in language, animating many of the most noticeable controversies in the philosophy of language: *descriptive* reference and *direct* reference. In a word, there are many ways, in terms of the nature and complexity of the relevant computations, in which vision 'anticipates' language or, to use the metaphor repeatedly proposed in this monograph, in which language 'mirrors' vision.

But isn't this eventually simply a useful metaphor? Well, judging on the grounds of the result of the most recent neuroscientific research, vision and language have much in common not only from the cognitive perspective adopted above, but also from a neuroanatomical perspective, in the way they are represented in the brain. In vision, information is transmitted from the primary visual cortex over two major pathways, a *ventral* pathway, leading from the occipital lobe into the temporal lobe, which encodes information about what the stimulus is, and a *dorsal* pathway, leading from the occipital into the parietal lobe, which encodes information about where the stimulus is and which is critical, as we have seen, for directing movement.

As for language, the relevant network includes the inferior frontal lobe, and more specifically Broca's area (with a functional distinction between B45, relevant for semantic processing, and B44, relevant for complex syntactic processing) and large portions of the temporal lobe, more specifically the *anterior temporal lobe* and the *posterior superior temporal lobe*. Interestingly, the language network is neuroanatomically represented in terms of *a ventral pathway* and a *dorsal pathway* (identified both in functional and cytoarchitectonic terms), each of which present a 'double' structure. Regarding the ventral pathway, Friederici (2017) proposes to

distinguish between: (a) the inferior-fronto-occipital fasciculus (IFOF), connecting B45/47 in the inferior frontal cortex to the temporal and occipital cortex, and (b) the fasciculus uncinatus, connecting the frontal operculum in the frontal cortex to the anterior temporal lobe. Regarding the dorsal pathway, the proposed distinction is between: (a) one dorsal fiber bundle connecting the temporal cortex to the premotor cortex via the parietal lobe, a pathway supporting auditory-to-motor mapping, and (b) the fasciculus arcuatus, connecting the posterior superior temporal lobe to B44, arguably recruited for the most complex modalities of syntactic processing. Research on language processing has insisted that the latter essentially correlates to a neurally represented capacity for analyzing the linguistic input in terms of hierarchical structures (Moro 2008, Friederici 2017). This suggests that the human capacity for syntax and syntactic processing is based on a phylogenetically determined ability to learn hierarchy-based grammars (mildly context-sensitive grammars), a sort of cognitive bias towards 'dendrophilia,' in Fitch (2014)'s sense.

Undoubtedly, however, the human capacity for syntax is also based on a phylogenetically determined ability to recruit pre-defined functional architectures, corresponding to a pre-established organization of the semantic/functorial space around the basic lexical categories (nouns, verbs, etc.). In fact, a large part of contemporary theoretical syntax consists in an in-depth investigation into the so-called functional structure of language. In its more general form, syntax puts together bundles of features by means of primitive structure-building operations. Lexical terms have a non-logical meaning and typically give rise to argument structure representations, which we might conceive of as the linguistic counterparts of complex visual scenes where participants take part in a unitary event of some sort. Functional words typically represent more abstract categories, such as noun class, gender, number, determiner, quantifier, case, verb class, voice, aspect, tense, modality, negation, degree, topic/focus, etc., and tend to express high semantic types, beyond the level of entities and properties (von Fintel and Matthewson 2008[30]). Whereas learning the open inventory of lexical items is influenced by experience and perception, the acquisition of this closed

---

30 See also the discussion in Chapter 3.

inventory of functional items is likely to represent a fixed property of the mind, in terms of a species-specific organization of the logical space, a property that apparently evolved into language in the form of complex templates of hierarchically ordered (bundles of) functional features, each expressing a logically sophisticated functorial operation (Cinque 1999, Borer 2005a,b). This unique organization of the conceptual/logical space in terms of pre-established templates of functional categories is a core property of language no less than dendrophilia. In fact, dendrophilia typically manifests itself in giving rise to complex pre-established hierarchical templates of functional categories around the basic syntactic categories.

The core role of both hierarchy and functional categories is confirmed by a rapidly converging amount of neurolinguistic evidence. More specifically, hierarchy-based long-distance dependencies have been shown to activate B44 with respect to non-hierarchical long-distance dependencies (Friederici and Bahlmann 2011). At the same time, sentences with pseudo-words involving inflectional morphology have been shown to activate B44, on a par with true sentences and in contrast to sentences that involve only derivational morphology (Goucha & Friederici 2015).

If functional structure is part and parcel of language as a formidable instrument for the expression of thought, the obvious question to be raised is how this level of computational complexity could evolve within a new single cognitive module. But this question – formulated for language – is not essentially different from the analogous questions that we may raise regarding the systems of perception. They are also characterized, as we have seen, by an astounding level of computational complexity (Gallistel 1996). The architecture of the mind does not reflect the evolution from simple cognitive systems (those responsible, according to the Cartesian view, for the passive registration of the sensory stimuli) to complex cognitive systems (those responsible, within the very same view, for the operations of thought and reason). Rather, each cognitive system is characterized by an astonishing level of computational complexity. And despite the functional and structural specificity of each cognitive system, what we arguably find is a significant amount of mirror effects, as if each of the systems could not only reproduce some of the features of the others, but also 'anticipate,' in *a priori* unpredictable ways, some of the features that are regarded as typical of another system.

It is in this vein that we have observed, in the preceding chapters, how the dichotomy between descriptive and direct reference, long-regarded as one of the defining features of the systems of interpretation in language, is 'anticipated' in vision, and in ways that are actually essential for a proper understanding of how vision works. It is in this vein that we have concluded, more generally, that the capacities for representation and reference arise primarily in perception, and are then 'reflected,' though with an increased formidable power, in language. It is in this vein that we have proposed that the semantics of the verbs of vision, as it can be reconstructed in a language such as English, somehow expresses and surprisingly 'anticipates' some of the ingredients of the epistemology of perception, as it independently takes shape as an effect of the human rational inquiry. And it is also in this vein that we have been led to propose that the syntactic templates associated to the verbs of propositional attitudes somehow encode the nature of such abstract entities as propositions.

These are simply some of the facts of language that keep startling us, as professional linguists. But in fact, the real 'take home message' is more general, it seems. If propositions, as expressed in language in the form of sentences, are information-extracting cognitive event types; if propositions, so conceived, are already expressed in perception before being expressed by means of linguistic tools; if perception is not only combinatory and compositional but also, and crucially, propositional; if sensory-features are wired-in classificatory devices and evolutionarily induced patterns of information extraction from analogical data bases (as it happens with retinotopic spatial relations in vision); if, finally, neurons themselves are computational machines performing a 'propositional' task; then the bewilderment we described above reveals its true source: the mind does not go from the extremely simple to the extremely complex; the mind is complex from the very beginning and the operations of the lower-level cognitive systems are never essentially 'passive' as the Cartesian tradition made us used to think.

Quite excitingly, this also seems to hold for the holy Cartesian pair of Thought and Reason. Mercier and Sperber 2017 take issue with Kahneman (2011)'s distinction between two modes of inference in thought, one occurring spontaneously and effortlessly (fast thinking) and the other occurring on the contrary deliberately and effortfully (slow thinking). In the Cartesian tradition, slow thinking is reasoning proper, it is what makes

us human. Reasoning entails conscious activation of explicit (and correct) mechanisms of logical inference. However, as we have seen, vision entails the activation of unconscious mechanisms of inference. Is human behavior dictated by explicit logic-based inference or by the implicit inferential devices that extensively undergo the activity of the different modules of the mind? Are we endowed with one Reason, encoded in slow thinking and typically expressed in language, or with many reasons, each of them obeying complex and essentially unconscious computational patterns? Is human decision-making based on the operations of Reason or more mundanely based on the complex unconscious interplay among the many operations of different kinds of reason?

Convincingly, Mercier and Sperber argue that Reason is most plausibly the product of the human ecological *niche*, in which decisions have to be defended and justified in a social setting, and people engage in logic-based reasoning only when there are social motivations for doing that, mainly the necessity of evaluating the others' suggestions and decisions, rejecting those with a weak foundation and only accepting those with a solid ground. In other words, the Aristotelian/Cartesian idea of solitary minds successfully engaged in correcting irrational and mistaken intuitions is fairly detached from reality. The alleged 'mistaken intuitions' are in fact the product of phylogenetically selected sophisticated computational modules. And an increasing body of empirical evidence suggests that corrections hardly take place, especially when they concern our own decisions. Logic – at least as it applies in Reason – and language are more the manifestation of the 'argumentative' compulsion originated by the social conditions of human life than of the presence, in humans, of a unique sort of higher cognition. Human decisions are commonly made as a result of the operations of unconscious – though computationally highly sophisticated – cognitive modules, on analogy with what happens with the other primates. Reason – when it applies – is positively biased towards the decisions we implicitly make and negatively biased towards the decisions that the others make. All in all, even from the perspective of practical rationality, the intellectual narcissism that we have identified above as one of the reasons for downplaying perception with respect to a higher system of 'active' operations of the mind (typically represented in language) is completely unfounded and essentially unjustifiable.

The feeling we get from all of this is that scrutinizing the mind is like scrutinizing a room full of looking-glasses. We are, unavoidably, prey of a sort of Borgesian feeling. And, as we already mentioned in one of the preceding chapters, we are led to wonder whether the looking-glass is not actually a rabbit-hole. Down the rabbit-hole then, down to a world where perception and cognition may continue to surprise us less for their differences than for what they have in common.

References

Aglioti, S., DeSouza, J. F. X., Goodale, M. A. (1995). "Size-contrast illusions deceive the eye but not the hand." *Current Biology*, 5, 679–685.

Akins, K. (1996). "Of Sensory Systems and the 'Aboutness' of Mental States," *Journal of Philosophy* 93: 337–72.

Anscombe, E. (1975). "The First Person," in S. Guttenplan (ed.), *Mind and Language: Wolfson College Lectures 1974*. Oxford: Oxford University Press, 45–64.

Armstrong, D.M. (1973). *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

Azzouni, J. (2013). *Semantic Perception. How the Illusion of a Common Language Arises and Persists*. Oxford: Oxford University Press.

Barsalou, L. W. (1999). "Perceptual symbol systems." *Behavioral and Brain Sciences*, 22(4), 577–660.

Barwise, J. (1982). "Scenes and Other Situations," *The Journal of Philosophy* 78, 369–397.

Bayer, J. (1986). "The role of event expression in grammar." *Studies in language* 10.1, 1–52

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

Borer, H. (2005a). *Structuring sense, vol. I: In name only*. Oxford: Oxford University Press.

Borer, H. (2005b). *Structuring sense, vol. II: The normal course of events*. Oxford: Oxford University Press.

Borer, H. (2013). *Structuring sense, vol. III: Taking form*. Oxford: Oxford University Press.

Botvinick, M., and Cohen, J. (1998). "Rubber hands 'feel' touch that eyes see." *Nature*, 391: 756.

Brogaard, B. (ed.) (2014). *Does Perception Have Content?* Oxford: Oxford University Press.

Brooks, L.R. (1978). "Nonanalytic concept formation and memory for instances." In Rosch E. and Lloyd B. (eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum, 169–211.

Budek, T. and Farkas, K. (2014). "Which causes of an experience are also objects of an experience?" In B. Brogaard (2014).

Burge, T. (1986). "Individualism and Psychology." *The Philosophical Review* 95 (1986), 3–45. Reprinted in *Foundations of Mind: Philosophical Essays*, Volume 2. Oxford: Clarendon Press, 2007.

Burge, T. (2005). "Disjunctivism and Perceptual Psychology." *Philosophical Topics* 33, 1–78.

Burge, T. (2010). *The Origins of Objectivity*. Clarendon Press, Oxford.

Capone, A. (2016). *The pragmatics of indirect reports: Sociophilosophical considerations*. Heidelberg: Springer.

Caramazza, A., and Mahon, B.Z. (2003). "The organization of conceptual knowledge: The evidence from category-specific semantic deficits." *Trends in Cognitive Sciences* 7, 354–361.

Caramazza, A., and Mahon, B.Z. (2006). "The organization of conceptual knowledge in the brain: The future's past and some future directions." *Cognitive Neuropsychology* 23, 13–38.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Castañeda, H. (1968). "On the Logic of Attributions of Self- Knowledge to Others." *Journal of Philosophy* 65:439– 56.

Cecchetto, C. and C. Donati (2015). *(Re)labeling*. Cambridge MA: MIT Press.

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

Cheng, L.L. and R.P.E. Sybesma (2012). "Classifiers and DP." *Linguistic Inquiry* 43(4): 634–650.

Chierchia, G. (1997). "Reference to kinds across languages." *Natural Language Semantics* 6: 339–405.

Chomsky, N. and Smith, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. New York: Oxford University Press.

Cinque, G. (2002) (ed.). *Functional Structure in DP and IP: The Cartography of Syntactic Structures*. Oxford: Oxford University Press.

Cohen, J. (2006). "Chromatic layering and color relationalism." *Minds and Machines* 26, 287–301.

Cohen, J. (2015a). "Perceptual integration, modularity and cognitive penetration." In Raftopolous, A. and Zembiek, J. (eds.), *Cognitive influences on perception: Implications for philosophy of mind, epistemology and philosophy of action*. Oxford: Oxford University Press, 123–143.

Cohen, J. (2015b). "Perceptual representation, veridicality and the interface theory of perception." *Psychonomic Bulletin & Review* 22, 1512–1518

Davidson, D. (2005). *Truth and Predication*. Cambridge, MA: Belknap Press.

Dehaene, S. (2014). *Consciousness and the brain. Deciphering how the brain codes our thoughts*. New York: Penguin Books.

Delfitto, D., A. Reboul and G. Fiorin (2017). "Immunity to error through misidentification and (direct and indirect) experience reports." In A. Capone, M. Garcia Carpintero, A. Falzone, eds. *Indirect reports (and pragmatics) in the world languages*. Cham, Springer.

Deroy, O. (2015). "Modularity of Perception." In M. Matthen (ed.), *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Dickie, I. (2015). "Perception and Demonstratives." In M. Matthen (ed.), *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Dretske, F.I. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F.I. (2000). "Simple seeing." In *Perception, Knowledge, and Belief: Selected Essays*. Cambridge: Cambridge University Press, 97–112.

Egan, F. (2010). "Computational models: a modest role for content." *Studies in History and Philosophy of Science* 41: 253–259.

Evans, G. (1982). *The varieties of Reference*. Oxford: Oxford University Press.

Fine, K. (2005). "Necessity and non-existence." In *Modality and tense: Philosophical papers*, Oxford: Oxford University Press, 321–353 (Chap. 9).

Fintel, K. and L. Matthewson (2008). "Universals in Semantics." *The Linguistic Review* 25, 139–201.

Fiorin, G. and Delfitto, D. (2020). *Beyond Meaning. A journey across Language, Perception and Experience*. Perspectives in Pragmatics, Philosophy and Psychology, vol. 25, Springer.

Fitch, T. (2010). *The Evolution of Language*. Cambridge and New York: Cambridge University Press.

Fitch, W.T. (2014). "Towards a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition." *Physics of Life Reviews*. 11: 329–364.

Fodor, J.A. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J.A. (2015). "Burge on Perception." In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.

Fodor, J. A., and Pylyshyn, Z. W. (2015). *Minds without Meanings: An Essay on the Content of Concepts*. Cambridge, MA: MIT Press.

Frege, G. (1918/19). "The Thought." In Brian McGuinness (ed.) *Collected Papers*. Oxford: Basil Blackwell, 1984.

Friederici, A. (2017). *Language in our brain. The origins of a uniquely human capacity*. Cambridge, MA: MIT Press.

Friederici, A. and J. Bahlmann (2011). "The neural basis of recursion and complex syntactic hierarchy." *Biolinguistics* 5.1–2, 87–104.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science* 4.1, 14–21.

Gallese V. and Lakoff G. (2005). "The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge." *Cognitive Neuropsychology* 22(3), 455–479.

Gallistel, C.R. (1996). "Insect Navigation: Brains as Symbol Processing Organs." In *Invitation to Cognitive Science*, iv, Cambridge, MA: MIT Press.

García-Carpintero, M. (2015). "De se thoughts and immunity to error through misidentification." *Synthese*, DOI: 10.1007/s11229-015-0817-y.

Gettier, E.L. (1963). "Is Justified True Belief Knowledge?" *Analysis*, 23(6): 121–123.

Gisborne, N. (2010). *The event structure of perception verbs*. Oxford: Oxford University Press.

Glüer, K. (2014). "Looks, Reasons and Experiences." In B. Brogaard (2014).

Goodale, M.A., and Milner, A.D. (1995), *The Visual Brain in Action*. Oxford: Oxford University Press.

Goodale, M.A., Pélisson, D., and Prablanc, C. (1986). "Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement." *Nature*, 320, 748–50.

Gopnik, A. and Schulz, L. (2007). *Causal Learning: Psychology, Philosophy and Computation*. Oxford: Oxford University Press.

Goucha, T. B., and Friederici, A.D. (2015). "The language skeleton after dissecting meaning: A functional segregation within Broca's area." *NeuroImage* 114 (6): 294–302.

Hauser, M., Chomsky, N., and Fitch, W. T. (2002). "The Faculty of Language: What is it, Who has it, and How did it Evolve?" *Science* 298, 1569–1579.

Hickok, G. (2014). *The myth of mirror neurons. The real neuroscience of communication and cognition*. Norton & Company, New-York/London.

Higginbotham, J. (1983). "The Logic of Perceptual Reports: An Extensional Alternative to Situation Semantics." *The Journal of Philosophy* 80.2, 100–127.

Higginbotham, J. (2003). "Remembering, Imagining, and the First-Person." In A. Barber (ed.), *Epistemology of Language*. Oxford: Oxford University Press.

Hintikka, J. (1998). "Perspectival Identification, Demonstratives and 'Small Worlds.'" *Synthese* 114, 203–232.

Hinzen, W. (2014). "What is Un-Cartesian Linguistics?" *Biolinguistics* 8, 226–257.

Hinzen, W. and Sheehan, M. (2013). *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.

Horgan, T.E., Tienson, J.L. and Graham, G. (2004). "Phenomenal intentionality and the brain in a vat." In R. Schantz (ed.), *The Externalist Challenge*. Berlin: De Gruyter.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jacob, P. and Jeannerod, M. (2003). *Ways of seeing: The scope and limits of visual cognition*. Oxford: Oxford University Press.

Kahneman, D. and Treisman, A. (1992). "The reviewing of object files: object-specific integration of information." *Cognitive Psychology* 24, 175–219.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss and Giroux.

Kandel, E. et al. (2013). *Principles of neural science*. Fifth edition. The McGraw-Hill Companies.

Kaplan, D. (1989). "Demonstratives." In Almog, J., Perry, J. and Wettstein, H. (eds.), *Themes from Kaplan*. Oxford: Oxford University Press.

Kaup, B., Zwaan, R. A., and Lüdtke, J. (2007). "The experiential view of language comprehension: How is negation represented?." In Schmalhofer, F. Perfetti, C. A. Mahwah (eds.), *Higher level language processes in the brain: Inference and comprehension processes*. NJ: Lawrence Erlbaum Associates.

Kayne, R. (2010). *Comparisons and contrasts*. Oxford Studies in Comparative Syntax. Oxford: Oxford University Press.

King, J., S. Soames, and J. Speaks (2014). "*New Thinking about Propositions*." Oxford: Oxford University Press.

Kiparsky, P., and C. Kiparsky (1970). "Fact." In M. Bierwisch and K. E. Heidolph (eds.), *Progress in Linguistics*. The Hague: Mouton.

Krifka, M. (2010). "A note on the asymmetry in the hedonic implicatures of olfactory and gustatory terms." In Susanne Fuchs, Phil Hoole, Christine Mooshammer & Marzena Zygis (eds.), *Between the regular and the particular in speech and language*, 235–245. Frankfurt/M.: Peter Lang.

Kripke, S.A. (1980). *Naming and Necessity*. Harvard: Harvard University Press.

Kripke, S.A. (2011). "The first-person." In *Philosophical Troubles. Collected Papers Vol I*. Oxford: Oxford University Press.

Lakoff, G. (1987). "Cognitive models and prototype theory." In U. Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge University Press.

Landau, B. and Jackendoff, R. (1993). " 'What' and 'where' in spatial language and spatial cognition." *Behavioral and brain sciences* 16.2: 217–238.

Langacker, R.W. (1987). *Foundations of cognitive grammar*, Volume I, Theoretical Prerequisites. Stanford, CA: Stanford University Press.

Longobardi, G. (1994). "Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form." *Linguistic Inquiry* 24: 4.

Lycan, W.G. (2014). What does Vision Represent? In Brogaard (2014).

Macdonald G.F. (ed.) (1979). *Perception and Identity: Essays Presented to A. J. Ayer, with His Replies*. London: Macmillan.

Machery, E. (2009). *Doing without Concepts*. Oxford: Oxford University Press.

Marr, D. (1982). *Vision*. New York: Freeman.

Matthen, M. (2005). *Seeing, doing and knowing. A philosophical theory of sense perception*. Oxford: Oxford University Press.

Matthen, M. (2014). "Image Content." In B. Brogaard (2014).

Matthen, M. (2015a). "Introduction." In M. Matthen (ed.), *The Oxford handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Matthen, M. (2015b). "The individuation of the senses." In M. Matthen (ed.), *The Oxford handbook of Philosophy of Perception*. Oxford: Oxford University Press.

May, R. (1991). "Syntax, semantics and logical form." In Asa Kasher (ed.), *The Chomskyan Turn*. Oxford: Blackwell.

McGurk, H. and MacDonald, J. (1976). "Hearing lips and seeing voices." *Nature* 264: 746–748.

Medin, D.L. and Schaffer, M.M. (1978). "Context theory of classification learning." *Psychological Review* 85: 2017–238.

Mercier, H. and Sperber, D. (2017). *The enigma of reason*. Harvard: Harvard University Press.

Millikan, R.G. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.

Milner, A.D. and Goodale, M.A. (2006). *The Visual Brain in Action*. Oxford Psychology Series: Vol. 43. Oxford: Oxford University Press.

Moeschler, J. (2014). "How logical are logical words?" In M. Taboada and M. Trnavac, *Non-veridicality and evaluation*. Studies in Pragmatics, Brill, 76–110.

Moltmann, F. (2013). *Abstract Objects and the Semantics of Natural Language*. Oxford: Oxford University Press.

Moltmann, F. (2016). "Natural language ontology." *Oxford Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Moro, A. (2008). *The boundaries of Babel. The brain and the enigma of impossible languages*. Cambridge, MA: MIT Press.

Moro, A. (2016). *Impossible languages*. Cambridge, MA: MIT Press.

Mulligan, K. (1999). "Perception, Particulars and Predicates." In Fisette D. (ed.), *Consciousness and Intentionality: Models and Modalities of Attribution*. Dordrecht: Springer.

Murphy, G.L. and Medin, D.L. (1985). "The role of theories in conceptual coherence." *Psychological Review* 92, 289–316.

Noë, Alva, and O'Regan, Kevin (2002). "On the Brain-Basis of Visual Consciousness: A Sensorimotor Account," in A. Noë and E. Thompson (eds.) (2002). *Vision and Mind: Selected Readings in the Philosophy of Perception*. Cambridge, Mass.: Bradford Books, MIT Press, 67–98.

Ogden, C.K. and Richards, I.A. (1923). *The Meaning of Meaning*. London: Routledge and Kegan Paul.

Parsons, T. (1972). *An Outline of Semantics of English*. Amherst: Ms. University of Massachusetts.

Peacocke, C. (1983). *Sense and Content: Experience, Thought and Their Relations*. Oxford: Oxford University Press.

Peacocke, C. (2005). "Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge," in N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds.), *Joint Attention: Communication and Other Minds*, Oxford: Oxford University Press.

Peacocke, C. (2015). *The mirror of the world. Subjects, consciousness and self-consciousness*. Oxford: Oxford University Press.

Perry, J. (1979). "The Problem of the Essential Indexical," *Noûs* 13, 3–21.

Piccinini, G. (2008). "Computation without representation." *Philosophical Studies*, 137: 205–241.

Pietroski, P. (2005). "Meaning Before Truth." In G. Preyer and G. Peter (eds.), *Contextualism in Philosophy*, 253–300. Oxford: Oxford University Press.

Pietroski, P. (2018). *Conjoining Meanings. Semantics without Truth-Values*. Oxford: Oxford University Press.

Pulvermüller, F. (1999). "Words in the brain's language." *Behavioral and Brain Sciences*, 22(2), 253–336.

Putnam, H. (1975). *The Meaning of 'Meaning.'* Minneapolis: University of Minnesota Press.

Pylyshyn, Z. (1986). *Computation and Cognition*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (1999) "Is vision contiguous with cognition? The case for cognitive impenetrability of visual perception." *Behavioral and Brain Sciences* 22, 341–423.

Pylyshin, Z.W. (2007). *Things and places. How the mind connects with the world*. Cambridge, MA: MIT Press.

Ramchand, G. (2018). *Situations and Syntactic Structures*. Cambridge, MA: MIT Press.

Reboul, A. (2017). "Cognition and communication in the evolution of language." *Oxford Studies in Biolinguistics*. Oxford: Oxford University Press.

Recanati, F. (2007). *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.

Recanati, F. (2012). "Immunity to error through misidentification: what it is and where it comes from." In Prosser, S. and F. Recanati (eds.), *Immunity to error through misidentification*. Cambridge: Cambridge University Press.

Reuland, E. (2011). *Anaphora and Language Design*. Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Roberts, I., and Roussou, A. (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Rosch, E. (1973). "On the internal structure of perceptual and semantic categories." In Moore, T. (ed.), *Cognitive Development and Acquisition of Language*. New York: Academic Press, 111–144.

Saito, M. (2015). "Notes on the referential transparency of perception and factive verb complements." *Nanzan linguistics* 10, 21–42.

Scappini, M., Vespignani, F., Savazzi, S., Marzi, C. and Delfitto, D. (2016). "On the non-incremental processing of negation: A pragmatically licensed sentence-picture verification study with Italian (dyslexic) adults." *Nouveaux cahiers de linguistique française* 32.

Schellenberg, S. (2014). "The Relational and Representational Character of Perceptual Experience." In B. Brogaard (2014).

Sellars, W.S. (1956). "Empiricism and the Philosophy of Mind." In Feigl, H. & Scriven M. (eds.), *Minnesota Studies in the Philosophy of Science*, vol. I, Minneapolis, MN: University of Minnesota Press, 253–329.

Shoemaker, S. (1968). "Self-reference and Self-awareness," *Journal of Philosophy*, 65/19: 555–567.

Shoemaker, S. (1994). "Phenomenal Character," *Noûs* 28 (1994): 21–38.

Shoemaker, S. (1996). *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.

Siegel, S. (2014). "Affordances and the Contents of Perception." In B. Brogaard (2014).

Simons, J. and Chabris, C.F. (1999). "Gorillas in our midst: sustained inattentional blindness for dynamic events." *Perception* 28, 1059–1074.

Soames, S. (2015). *Rethinking Language, Mind and Meaning*. Princeton: Princeton University Press.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge, MA: MIT Press.

Tenny, C. L. (2006). "Evidentiality, Experiencers and the syntax of sentience in Japanese." *Journal of East Asian Linguistics*, 15:245.

Tye, M. (2014). "What is the Content of a Hallucinatory experience?" In B. Brogaard (2014).

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Vender, M. and D. Delfitto (2010). "Towards a pragmatics of negation: The interpretation of negative sentences in developmental dyslexia." *GG@G – Generative Grammar at Geneva* 6:1–28.

Vulchanova, M., Vulkanov, V., Fritz, I. and Milburn, E.A. (2019). "Language and perception: Introduction to the Special Issue 'Speakers and Listeners in the Visual World.'" *Journal of Cultural Cognitive studies* 3.

Weiskrantz, L. (1986). *Blindsight. A case study and implications*. London: Oxford University Press.

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

Wright, W. (2015). "Nonconceptual Content." In M. Matthen (ed.), *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Index of Names

Studies in Philosophy of Language and Linguistics

Edited by Piotr Stalmaszczyk

www.peterlang.com