

# 1 The representational theory of the visual mind

---

In the present chapter, we sketch and argue for a view, which we call the ‘representational theory of the visual mind’ (RTVM). RTVM is not so much a scientific theory that leads to testable predictions, as a picture or a framework. According to the representational theory of the mind, the mind is at bottom a representational device: in Dretske’s (1995b: xiv) terms, ‘all mental facts are representational facts’. On this view, mental processes consist of the formation and the transformation of mental representations.

In Section 2 of the present chapter, we shall contrast our version of RTVM (which we call ‘visual intentionalism’) with two alternatives: ‘sense-datum theory’ and ‘disjunctivism’, the latter of which is advocated by some contemporary ‘direct realists’. Visual intentionalism, as we conceive it, will turn out to offer a middle course between sense-datum theory and disjunctivism. At the end of Section 2, we shall face the challenge that any representational approach must face, namely the challenge of the homunculus. In our view, RTVM is a framework for thinking about two main puzzles: the puzzle of the visual perception of objects and the puzzle of object-directed actions.

The puzzle of visual perception is the puzzle of how a purely subjective visual experience can provide us with objective knowledge of the world. This puzzle will be taken up again in more detail in Chapter 5. In the present chapter, we call attention to two features of visual percepts. First, in Section 3, we sketch our reasons for thinking that visual percepts have non-conceptual content: we examine the paradigmatic arguments from philosophers who appeal to the distinctive phenomenology of visual experience in order to justify the distinction between the conceptual content of thoughts and the non-conceptual content of visual experiences. Second, in Section 4, we sketch the basis of an approach that we label ‘cognitive dynamics’, whose purpose is to provide a systematic understanding of the dynamical mapping from visual percepts to thoughts and from more ‘engaged’ to more ‘detached’ thoughts. Thus, much of the present chapter is a detailed exploration of the resources of RTVM. One goal of Section 4 on cognitive dynamics is to show that RTVM is not committed to the view that all mental representations are detached descriptive concepts. Not all mental representations need have purely conceptual descriptive content.

The puzzle of visually guided actions is the puzzle of how so many human actions directed towards a target can be accurate in the absence of the agent’s visual awareness of many of the target’s visual attributes. In Section 5, we turn our attention to three implications of RTVM for the control of visually guided actions. We shall argue that RTVM has the resources to clarify the puzzle of visually guided actions. First, we examine the nature of actions and argue that actions involve mental representations. In Chapter 6, we shall further characterize the specific content of ‘visuomotor’ representations. Second, we

examine the ineliminable role of intentions in the etiology of actions. Third, we discuss the intentionality of intentions. We argue that what is distinctive of intentions is that they have a world-to-mind direction of fit, and a mind-to-world direction of causation. This combination explains the peculiar commitment of intentions to action. In Chapter 6, we shall rely on these ingredients of RTVM to argue that there is a basic asymmetry between visual percepts and visuomotor representations. While the former is input to a process whose output is stored in the ‘belief box’, the latter is at the service of the ‘intention box’.

## 1 A teleosemantic account of visual percepts

Tokens of mental representations are best thought of as tokens of an animal’s brain states or states of its central nervous system. Not any internal physiological state of an animal, however, is a mental representation. States of an animal’s digestive system, of its cardio-vascular system, of its respiratory system or of its immune system are not mental representations. Mental representations are neurophysiological states with *content*. Nor are all representations mental representations. The evolution of human cognition has given rise to cultural artifacts, i.e. to non-mental representations of various sorts, such as linguistic utterances, mathematical and logical symbols, diagrams, road-signs, maps, states of measuring devices (e.g. gauges, thermometers, scales, altimeters, etc.), paintings, drawings, photographs and movies. Thoughts, judgments, beliefs, desires, intentions, perceptual experiences, memories and mental images are mental representations. Whether mental or non-mental, all representations have content. They may also have computational properties: as emphasized by many philosophers and cognitive scientists, mental and non-mental representations are typically things to which computations apply and which can be studied from a computational point of view.<sup>1</sup>

We assume, along with many philosophers, that artifacts (i.e. non-mental representations) derive their contents from the contents of the mental representations of the human beings who create and/or use them. We therefore subscribe to the distinction between the *primitive* intentionality of mental representations and the *derived* intentionality of artifacts.<sup>2</sup> Although artifacts derive their contents from the primitive contents of the mental representations of their creators and users, unlike mental representations, they are publically observable. Non-mental representations are physical structures, and as such they have intrinsic physical and chemical properties. What makes them representations is that they have contents. In our view (much inspired by Dretske 1988, 1995b), representations are physical structures with informational function, i.e. with the function to carry information.<sup>3</sup>

A physical signal  $S$  can be said to carry information about property  $F$  if  $S$  tracks instances of  $F$  or if  $S$  is reliably (or nomically) correlated with exemplifications of  $F$ .

<sup>1</sup>See Fodor (1975, 1987), Chomsky (1980, 2000), Marr (1982), Pylyshyn (1984), Peacocke (1994).

<sup>2</sup>The distinction is accepted by Fodor (1987), Searle (1992), Dretske (1995b) among others. It is rejected by Dennett (1987).

<sup>3</sup>For elaboration, see Jacob (1997).

Thus, the informational relation between  $S$  and  $F$  is taken to be the converse of the correlation between  $F$  and  $S$ . For example, tracks, fingerprints, states of measuring devices and symptoms all carry information. Information so conceived is what Grice (1957) called ‘natural meaning’. Perhaps a signal that carries information about a property could be called a ‘natural sign’. A track in the mud carries information about the kind and the size of the animal that left it. A fingerprint carries information about the identity of the human being whose finger was imprinted. A gas-gauge on the dashboard of a car carries information about the amount of fuel in the car tank. Spots on a human face carry information about a disease. In all such cases, a signal carries information about some property because the activated signal is correlated with the property in question, and the correlation is reliable and not purely accidental. Informational semantics (i.e. semantics based on non-coincidental correlations) is an essential tool of cognitive neuroscience. Cognitive neuroscientists try to map the activation of neurons in selected areas of the visual system with particular visual attributes instantiated by objects in the environment. To discover that neurons fire in response to (or ‘code’) the presence of a visual attribute, is to discover that the pattern of neuronal discharge carries information about the attribute in question. Thus, the length of a simple metal bar carries information about the temperature because the length of the metal bar nomically covaries with the variations of the temperature. If so, then the metal bar is a reliable indicator of the temperature. For two connected reasons, information so defined falls short of representation: on the one hand, information is ubiquitous; on the other hand, natural signs cannot misrepresent.

First, if  $S$  carries information about  $F$  and  $F$  is correlated with  $G$ , then  $S$  carries information about  $G$ . The informational relation being transitive, information is ubiquitous and, unlike semantic content, informational content is indeterminate. For example, the length of the metal bar carries information about the temperature. But if variations in temperature are in turn correlated with variations in atmospheric pressure, then the length of the metal bar carries information about atmospheric pressure. Representing the temperature, however, is not representing atmospheric pressure. Hence, given that the length of the metal bar carries information about both the temperature and atmospheric pressure, it cannot *represent* the temperature at all. Similarly, there are many stages on the way from the retina through the optic nerve to the higher levels of information-processing in the visual cortex. Each such stage carries some information about the distal stimulus and about everything the distal stimulus stands in some non-accidental correlation with. However, neither the retina nor the optic nerve represent everything they carry information about.

Second, unless a signal could *misrepresent* what it indicates, it cannot represent it. Unlike mental and non-mental representations, natural signs cannot fail to mean or carry information. As Dretske (1988: 56) puts it, ‘a person can *say*, and *mean*, that a quail was here without a quail’s having been here. But the tracks in the snow cannot mean (in the natural sense of ‘meaning’) that a quail was here unless, in fact, a quail *was* here’. Unlike a metal bar, a mercury thermometer may represent the temperature. What is the difference? Unlike a simple metal bar, a mercury thermometer does misrepresent the temperature if it malfunctions or if it does not work according to its design. For example, if the glass pipe containing the mercury is broken, then the thermometer may malfunction and misrepresent the temperature. Similarly, the gas-gauge in a car is

a representational system whose function is to indicate the amount of gas in the car's tank. Since it has the function to carry information about the amount of fuel in the tank, it too can malfunction and therefore misrepresent how much gas is left in the tank. First, as Fodor (1987) has put it, no representation without *misrepresentation*. Second, 'teleosemantic' theories add: no misrepresentation without a *function*. At the heart of the teleosemantic conception of content is the claim that a representational device owes its content to what Millikan (1984, 1993) calls the device's 'proper' function, where proper function is a teleological, not a dispositional notion. Third, a device's proper function derives from the device's *history*.

Arguably, unless a device has a function, it makes no sense to say that it is malfunctioning. Unless it has a function, a device cannot be defective, damaged or dysfunctional. Thus, unless its function is to carry information about some property, a device cannot be said to misrepresent the exemplification of the property. Unless a device has the function to indicate the temperature, it cannot misrepresent the temperature. A microphone is an electro-acoustical device whose function is to convert the energy of acoustic waves into electrical energy. So is the human ear. They both contain a diaphragm that responds to acoustic vibrations. Unless they had this function, they could not fail to transmit information about sounds. Hence, they could not be said to represent sounds.

Arguably, nothing can have a function unless it has a history. More precisely, nothing can have a function unless it results from some historical process of selection. The historical process of selection is the source of the device's design. Selection processes are design processes. Thus, according to 'teleosemantic' theories, design is the main source of function and content depends on informational function. Such theories are called teleosemantic theories in virtue of the connection between design or teleology and content.

Now, selection processes can be intentional or non-intentional. Mental representations derive their informational functions from a *non*-intentional selection process. The paradigmatic non-intentional process is the mechanism of natural selection by which Darwin explained the phylogenetic evolution of species: natural selection sorts organisms that survive, but no intentional agent is responsible for the sorting. The process of natural selection, is, as Kitcher (1993) puts it, a design process 'without a designer'. The sensory mechanisms of human and non-human animals have informational functions: the visual system, the auditory system, the olfactory system, the tactile system are complex biological systems. They have been recruited by natural selection because they carry information about different specific sets of properties that were instantiated in the environment of human ancestors and early humans in the course of evolution.

In fact, according to the so-called 'etioloical' theory of functions—argued for by Wright (1973) and defended by many teleosemanticists such as Millikan (1984, 1993) and Neander (1995)—functions are selected effects. The function or functions of a device must be effects of the device: they must be things that the device can do. What a representational device represents depends on its informational functions. Its informational functions in turn depend on what properties the device can carry information about. The properties a device carries information about are properties the device is nomically correlated with.

An artifact containing a column of mercury responds to pressure. Knowing this, we can use such a device to represent variations in altitude. But variations in the height of

a column of mercury are correlated with variations in atmospheric pressure. Which properties an animal's sensory system responds to is not up for us to decide.<sup>4</sup>

Since it cannot reliably discriminate between flies and the movements of lead pellets, the frog's visual system represents small black moving dots, not flies. Frogs feed on flies, not on lead pellets, but they catch flies by means of the visual representation of small black moving dots. The point was made by Fodor (1987: 131–2) in relation to the perceptual environment of male sticklebacks. Male sticklebacks detect sexually active male competing sticklebacks by their characteristic red spot. Upon detecting the characteristic red spot on a sexually active male stickleback, another male stickleback will respond by a no less characteristic display of territorial behavior. But, as Fodor (*ibid.*) puts it, 'the stupidity of the whole arrangement is immediately manifest when an experimenter introduces an arbitrary red object into the scene. It turns out that practically anything red elicits the territorial display; a breeding stickleback male will take Santa Claus for a rival'. The visual system of male sticklebacks represents the presence of red spots, not the presence of other sexually active male sticklebacks.

The sand scorpion is a nocturnal animal: at night, it emerges from its burrow to feed and to mate. It feeds on anything that it can hold onto long enough to paralyze with its neurotoxic sting located at the end of its tail. As discussed by Brownell (1984), it lacks sophisticated visual, auditory and olfactory detection mechanisms: 'covering all eight of the animal's eyes with opaque paint had no effect on either the scorpion's sensitivity to threatening stimuli or on the accuracy with which it turned toward them. Inserting sound-absorbent tiles between the stimulus and the scorpion also did not affect its responses'. A moth held squirming in the air a few centimeters from a scorpion fails to attract its attention. Brownell (1984) reports that the sand scorpion has tarsal hairs and basitarsal slit sensilla at the end of its legs, whose sensory neurons detect vibrations produced in the sand by either prey or predators. Insects cause vibrations in the sand. But so do gentle disturbances of the sand intentionally produced with a twig by an experimental scorpion-psychologist. Vibrations in the sand produced by the motion of a twig do trigger a scorpion's attack. The sensory mechanisms available to the sand scorpion do not allow it to discriminate the vibrations produced by an insect from those produced by a twig. Although the sand scorpion feeds on insects, not on twigs, nonetheless what the sand scorpion's receptors represent are vibrations in the sand, not the insects that cause them.

Dolphins are known to have a sonar system that is sensitive to the geometric shapes of objects. Suppose with Dretske (1990a) that we train a dolphin to discriminate shapes exhibited in water. Suppose that the dolphin learns to recognize all, and only, cylindrical objects in the water. Suppose, further, that all and only the cylindrical objects that have been included in the sample to which the dolphin has been exposed are made in plastic. The dolphin has learned to discriminate cylindrical objects and although all the cylindrical objects that the dolphin is able to recognize are made of plastic, still the dolphin has not learned to recognize plastic objects. Why? Simply because the sensory

<sup>4</sup>Some philosophers, e.g. Dennett (1987) and Searle (1992), disagree and argue that there is no fact of the matter as to what the function(s) is of a biological organ.

mechanism that allows the dolphin to recognize shapes is a sonar system. This sensory mechanism is sensitive to the shape, not to the chemical structure of objects.<sup>5</sup>

All four examples—the frog, the stickleback, the sand scorpion and the dolphin—show the need for careful investigation of the sensitivity of an animal’s sensory *mechanisms*. It is not enough to know what a predator feeds on in order to know how its sensory system represents its prey. Property *G* matters to the survival of the animal (e.g. a sexually active male competitor or an insect to capture). The animal’s sensory mechanism, however, responds to instantiations of property *F*, not property *G*. Often enough in the animal’s ecology, instantiations of *F* coincide with instantiations of *G*. So detecting an *F* is a good cue if what enhances the animal’s fitness is to produce a behavioral response in the presence of a *G*. But the animal does not represent *G* as such. The correlational or informational part of the teleosemantic account of mental representations is precisely designed to take into account the capacities of the sensory mechanisms.<sup>6</sup>

Only if a system is tuned to respond reliably to instantiations of *F* will it be able to tell if *F* is being instantiated. In fact, as we said above, the correlational or informational component of the teleosemantic approach underlies the practice of much cognitive neuroscience, whose project is to map the electrophysiological activity of some selected brain area onto the instantiation of some specific property. As we said above, when cognitive neuroscientists speak of the pattern of neural discharge as ‘coding’ for a given property, they rely on a correlational or informational relation between some brain area and the exemplification of a given property in the brain’s environment. Reliability, however, does not mean infallibility: misfiring may occur at some stage in the system.

Thus, the primate visual system evolved because it had the ability to carry information about the size, shape, orientation, internal structure, contours, texture, color, spatial position, distance and motion of objects. Ancestors of humans with such visual abilities survived in the competition against creatures with different visual abilities. As a result of natural selection, the human visual system has acquired the biological function to carry information about such properties. As much contemporary cognitive neuroscience of vision has taught us (see, e.g. Zeki 1993), different visual attributes of objects are processed in separate cortical areas in the visual brain of primates: neurons in area V3 respond to moving shapes; neurons in area V4 respond to colors; neurons in areas MT and V5 are specialized for the processing of motion. Each of these distinct brain areas has been shaped by evolution and selected for responding to specific

<sup>5</sup>Here we are claiming that the contents of the representations delivered by such a sensory system as a frog’s visual system or a dolphin’s sonar system are sharply constrained by the psychophysics of the sensory mechanisms. As we shall see in Chapter 5, although the contents of human visual perceptual representations too are constrained by the kinds of properties (size, shape, texture, orientation, color, etc.) that the human visual system has been designed to process, it does not follow that humans cannot visually represent the cheerfulness, sadness, threat or anger of a conspecific’s face. No doubt one can come to believe through visual perception that a face is cheerful, sad, threatening or angry. One can visually represent a face *as* (or see *that* a face is) cheerful, sad, threatening or angry (see Chapter 7). Humans can do this because they have the resources to move from non-epistemic to epistemic visual perception (see Introduction and Chapter 5).

<sup>6</sup>According to *pure* teleological views, such as Millikan’s (1984, 1993), the animal’s sensory mechanism represents what enhances the animal’s overall fitness, i.e. property *G*, not *F*. By not taking into account the mechanisms involved in the production of sensory representations, such accounts adopt an exaggerated version of adaptationism. See Jacob (1997, 2001) for more details.

visual attributes. As a result of a lesion in one area of the visual system, the visual system may fail to perform one of its particular informational functions: it may fail to carry reliable information about the shape, color, texture, position or motion of objects. As a result of a lesion in a highly specific brain area, a human patient will fail to experience, e.g. color in the case of achromatopsia, shape in the case of visual-form agnosia, motion in the case of akinetopsia.

According to RTVM then, the phenomenal qualities of an experience are the properties that objects are represented as having in the experience. The phenomenal properties of a visual experience are the ‘intentional’ properties the visual stimulus is represented as exhibiting by the experience itself. Visual experiences have a distinctive phenomenology different from the phenomenology of experiences in different modalities because the human visual system has been selected to respond to basic properties of objects that are different from the basic properties of objects to which the other human sensory systems have been selected to respond. Visual perception makes us aware of such fundamental properties of objects as their size, orientation, shape, color, texture, spatial position, distance and motion, all at once. So colors can be seen but they cannot be smelled, heard or touched. By contrast, sounds can be heard but they cannot be seen. Pressure can be felt but it cannot be seen either.<sup>7</sup>

What are crucial to *visual* phenomenology are those attributes of objects that can be processed visually and not otherwise (i.e. not by audition, smell or touch). One and the same object (e.g. a violin) can exemplify properties that can be processed in different modalities. Obviously, one thing is to see a violin. Something else is to hear the sound of a violin.

Now, the question arises: are there not properties of objects that can be processed in more than one sensory modality? For example, the shape of an object can be seen and it can also be touched or felt. Nonetheless, it might be objected, seeing the shape of a cube and touching it are very different phenomenal experiences. What it is like to see a cube is clearly different from what it is like to touch it. If so, then does it not follow that the phenomenology of sensory experience cannot be identified with the property the object is represented as having in the experience? No, it does not because there is indeed a difference between the way vision and touch present the shape of a cube. A normally sighted person will not see the shape of the cube without seeing its color. But by feeling the shape of a cube, one does not thereby feel its color. So although the shape of an object can be both seen and felt, still the phenomenal experience of seeing the shape differs from the phenomenal experience of feeling it because only the former will reveal the color of the object whose shape is being seen.<sup>8</sup>

Hence, the difference in the phenomenal character of seeing a shape and feeling it can be made to square with the representational view of the visual mind: the difference in phenomenal character arises from a difference between the visual and the tactile representation of the shape. Indeed, although the property represented by vision and by touch might be the same, the visual perceptual mode of perceiving shape differs from the tactile perceptual mode of perceiving it.

<sup>7</sup>As it will turn out, in order to capture the fine-grained non-conceptual content of a visual perceptual representation, which property is being represented will not suffice. We shall introduce the notion of a visual perceptual mode of presentation.

<sup>8</sup>See Chapter 5 for an extended discussion.

## 2 Visual intentionalism, sense-data and disjunctivism

According to RTVM, visual perception consists in forming and transforming mental representations. Now, the appeal to mental representations is traditional in the philosophy of visual perception: it is at the core of ‘sense-datum’ theories. Conversely, several contemporary philosophers, who subscribe to direct realism, have expressed scepticism towards the appeal to mental representations in cognitive science. Thus, Putnam (1994: 453) writes:

[...] in contemporary cognitive science, for example, it is the fashion to hypothesize the existence of ‘representations’ in the cerebral computer. If one assumes that the mind is an organ, and one goes on to identify the mind with the brain, it will become irresistible to (1) think of some of the ‘representations’ as analogous to the classical theorist’s ‘impressions’ (the cerebral computer, or mind, makes inferences from at least some of the ‘representations’, the outputs of the perceptual processes, just as the mind makes inferences from impressions, on the classical story), and (2) to think that those ‘representations’ are linked to objects in the organism’s environment only causally, and not cognitively (just as impressions were linked to ‘external objects’ only causally, and not cognitively).

At one extreme of the spectrum of views in the philosophy of perception, lie sense-datum theories. At the other extreme lie direct realist views. According to the former, visual perception consists in being aware of visual sense-data. Sense-data are mental ‘impressions’ that bear the properties one is aware of in visual perception. The latter embrace a radical form of externalism according to which we should give up the very idea of an ‘interface’ between the mind and the world. But, as it will turn out, the price to pay for giving up the idea of an interface between the mind and the world seems to be that the world itself turns out to be mind-dependent. In this section, we want to examine precisely the respects in which RTVM—or visual intentionalism—differs from both of these extreme views.

### 2.1 Sense-data and the argument from illusion

Visual perception gives rise to subjective experiences with a peculiar phenomenal character and it yields objective knowledge of the world. It is not surprising therefore that issues of visual phenomenology have been, and still are, intertwined with epistemological issues in the philosophy of visual perception. The epistemological goal of much traditional philosophy of perception has been to locate a secure foundation upon which to erect the rest of human knowledge. Many philosophers have assigned this foundational epistemological role to the concept of a *sense-datum*. Thus, Russell (1911) famously distinguished between ‘knowledge by acquaintance’ and ‘knowledge by description’. Since one can be acquainted with individuals or particulars, knowledge by acquaintance is non-propositional knowledge of objects. Unlike knowledge by acquaintance, knowledge by description is propositional knowledge of facts about objects. Thus, being simpler, knowledge by acquaintance is epistemologically prior to knowledge by description. The latter depends or supervenes on the former. According to Russell, however, genuine knowledge by acquaintance is not knowledge of ordinary physical objects: it is knowledge of mind-dependent or mental entities called ‘sense-data’. As Russell (1911: 73) wrote:

[...] in the presence of my table I am acquainted with the sense-data that make up the appearance of my table—its color, shape, hardness, smoothness, etc.; all these are things of which I am



immediately conscious when I am seeing and touching my table. The particular shade of color that I am seeing may have many things said about it—I may say that it is brown, that it is rather dark, and so on. But such statements, though they make me know truths *about* the color, do not make me know the color itself any better than I did before; so far as concerns knowledge of the color itself as opposed to knowledge about truths about it, I know the color perfectly and completely when I see it, and no further knowledge of it itself is even theoretically possible. Thus, the sense-data which make up the appearance of my table are things with which I have acquaintance, things immediately known to me just as they are.

On Russell's view then, visual sense-data are mental (or mind-dependent) entities. Unlike mind-independent physical objects, they can be directly known by introspection and with full Cartesian certainty. The mind is acquainted with nothing as fully and intimately as it is with itself. Visual sense-data are mental particulars and they have properties such as color and shape. So each of us is directly acquainted with one's visual sense-data and their properties. On the one hand, knowledge of truths about sense-data is indirect and depends on the more primitive introspective non-propositional acquaintance with them. On the other hand, propositional knowledge about mind-independent physical objects is achieved, if at all, by inference from knowledge of truths about sense-data. On Russell's view of acquaintance, the mind cannot be acquainted with mind-independent physical objects at all. Knowledge of, or about, mind-independent physical objects can only be knowledge by description, i.e. propositional knowledge. Knowledge of, or about, mind-independent objects is thus twice indirect: it derives from knowledge of truths about sense-data, which in turn depends on our prior acquaintance with sense-data. The chief epistemological motivation for postulating such mind-dependent entities as sense-data is that the mind can be directly acquainted with them and the process of acquaintance cannot go wrong. Acquaintance with mental entities provides an epistemically secure (though private and non-propositional) foundation upon which to erect the rest of human knowledge about the non-mental world.

Philosophers of perception, however, have had a second convergent motivation for postulating sense-data. As we pointed out in the previous section, only a device that may fail to give rise to veridical representations deserves to be called a representational system. Much traditional philosophy of perception has traded on the fact that the human perceptual system does not provide infallible knowledge of mind-independent objects. Sense-datum theory postulates that knowledge of sense-data is infallible. But perceptual knowledge of mind-independent objects is not. Thus, sense-datum theories have exploited the so-called 'argument from illusion', which, as we shall explain, is misleadingly so-called.

From a subjective point of view, the visual phenomenal appearances may perfectly well, it is claimed, be indistinguishable, whether the visual perception of mind-independent objects is veridical or not. Whether our visual experience of a non-mental object is veridical or not, there is something it is like to have it: something goes on in our minds in both veridical and non-veridical visual experiences. Something, therefore, is 'present to our minds' in both veridical and non-veridical visual perception. Since the visual appearances may be indistinguishable in both veridical and non-veridical visual perception, what is present to our minds, it is argued, must be common to veridical and to non-veridical cases of visual perception. Given that in non-veridical perception, it may be the case that no mind-independent entity is presented to the mind, it follows that

what is present to the mind in both veridical and non-veridical cases of visual perception is a mental sense-datum.

As Austin (1962) has pointed out in his devastating criticism of Ayer's (1940) version of the argument from illusion, much of its force depends upon a confusion between two quite distinct kinds of misperception: visual *illusions* and visual *hallucinations* (or as Austin calls them 'delusions'). As it will turn out in Chapter 4, in some circumstances, normally sighted people undergo size-contrast illusions such as the Ponzo illusion, the Müller-Lyer illusion or the Titchener circles illusion. Every normally sighted human being does. As we shall see in Chapter 4, size-contrast illusions arise from the attempt on the part of the visual perceptual system to maintain size constancy across a visual display containing elements of various relative sizes. In Austin's words:

[...] when I see an optical illusion [...] the illusion is not a little (or a large) peculiarity or idiosyncrasy of my own; it is quite public, anyone can see it, and in many cases, standard procedures can be laid down for producing it.

The argument from illusion, which would be better called 'the argument from delusion', can only go through if visual illusions are *delusive*, i.e. if as Austin (1962: 23–5) puts it, having a visual illusion consists in 'conjuring up' something 'immaterial'.

In fact, not only does the argument from illusion seem to involve a confusion between visual illusions and delusions, but it seems committed to subsuming under the category of illusions something that is not an illusion at all, namely seeing one's reflection in a mirror. From the fact that one sees one's face *in* a mirror, it does not follow, as Austin (*ibid.*, 31) notes, that one's face is actually located either in or behind the mirror. A proponent of the sense-datum theory would argue that what one sees then is a sense-datum. Following Tye (1995: 111–12), we would rather argue that this is evidence in favor of RTVM, i.e. that visual perception is *representing*. From the fact that one has a pain in one's left toe, and from the fact that one's left toe is in one's left shoe, it does not follow that the pain is in one's left shoe (or that there is a pain in one's left shoe). Nor does this show that the English preposition 'in' is ambiguous between a spatial and a non-spatial meaning. What it shows rather is that there is, as Tye (1995: 12) puts it, 'a hidden intensionality in statements of pain location'—as talk of pain in phantoms limbs confirms.

In this respect, visual experiences are like beliefs: they are mental representations. All representations are, in Quine's (1953) terms, *intensional* or *referentially opaque*. There are two criteria for referential opacity or intensionality. First, in belief contexts, co-referential expressions are not always substitutable *salva veritate*. Thus, one can believe that Cicero is bald and fail to believe that Tully is bald, even though 'Cicero' and 'Tully' are names of one and the same individual. Second, the rule of existential generalization does not always apply to beliefs: from the fact that someone believes that there are unicorns, it does not follow that there is any unicorn. Similarly, one can have a pain in one's left hand even though one's left hand has been amputated. On Tye's (1995) representationalist account, a phantom limb pain in one's amputated left hand is a mental *representation* of one's left hand. There need not be a left hand for one to represent it. Similarly, seeing one's face in the mirror is evidence that visual perception is forming a visual representation of one's face. From the fact that one sees one's face

in a mirror, it does not follow that one's face is located inside the mirror. That there is 'hidden intensionality' in reports of visual experiences, therefore, argues in favor of a representational view of visual experiences.

The argument from illusion starts from a standard case of a visual illusion, e.g. seeing a straight stick partly immersed in water as being bent. It then raises the puzzle of how something could be both 'crooked' and 'straight' without really changing its shape. Finally, it reaches the conclusion that 'at least some of the visual appearances are delusive'. As Austin (1962: 29) incisively writes:

[...] of this case Ayer says (a) that since the stick looks bent but is straight, 'at least one of the visual appearances of the stick is *delusive*'; and (b) that 'what we see [directly anyway] is not the real quality of [...] a material thing'. Well now: does the stick 'look bent' to begin with? I think we can agree that it does, we have no better way of describing it. But of course it does not look exactly like a bent stick, a bent stick out of water—at most, it may be said to look rather like a bent stick partly immersed *in* water. After all, we can't help seeing the water the stick is partly immersed in. So exactly what in this case is supposed to be *delusive*? What is wrong, what is even faintly surprising, in the idea of a stick's being straight but looking bent sometimes? Does anyone suppose that if something is straight, then it jolly well has to *look* straight at all times and in all circumstances? Obviously no one seriously supposes this.

The first crucial assumption in the argument from illusion is that veridical visual perception is the perception of 'material things' or, as Austin (*ibid.*: 8) calls them, 'moderate-sized specimens of dry goods'. The first assumption is that unless one sees a 'material thing', the visual appearances must be misleading or deceptive: the alternative is between the veridical seeing of 'material things' and the deceptive seeing of 'immaterial' (or 'unreal') ones. Either visual perception is veridical or it is not. If the former, then it is of 'material things'. If the latter, then it is of 'immaterial' or 'unreal' things. Thus, what Austin calls the 'bogus dichotomy' between the veridical perception of material things and its alleged alternative is supposed to prepare the ground for the second step in the argument. Whether they count as 'material things' or not, one can see rivers, substances, gases, vapors, mountains, flames, clouds, smoke, shadows, holes, pictures, movies and arguably events.<sup>9</sup>

The second step in the 'argument from illusion' trades on the confusion between visual illusions and visual hallucinations or delusions. Unlike seeing a straight stick partly immersed in water as being bent, seeing pink rats is suffering from a delusion. As Austin (*ibid.*: 23) argues, delusions are entirely different from visual illusions in that they involve high-level conceptual cognitive processes:

Typical cases would be delusions of persecution, delusions of grandeur. These are primarily a matter of grossly disordered beliefs (and so, probably, behavior) and may well have nothing in particular to do with perception. But I think we might also say that the patient who sees pink rats has (suffers from) delusions—particularly, no doubt, if, as would probably be the case, he is clearly aware that his pink rats aren't real rats.

Unlike visual illusions, which are pure perceptual processes and depend on perceptual constancy mechanisms, hallucinations involve (conceptual) belief-forming mechanisms.

<sup>9</sup>See Austin's (1962: 8) list.

As Dennett (1991) has noticed, ‘reports of very strong hallucinations are rare’. Phantom-limb pains are genuine cases of hallucination. But by Dennett’s (1991: 8) lights, they are weak, since they come in a single sensory modality: amputees *feel* their phantom-limbs, but they do not see, hear or smell them. Instances of genuine visual hallucinations, let alone multi-modal ones, are harder to come by than traditional philosophers of perception have been prone to assume.<sup>10</sup>

According to Dennett’s (1991: 8–17) model, hallucinations involve a lowering of the subject’s epistemic threshold for gullibility. For some reason (e.g. sensory deprivation, acute pain, extreme fear or trauma), subjects may lower their epistemic standards and become epistemically ‘passive’. As a result, ‘they feel no desire to probe, challenge or query’ the incoming information. Instead, ‘they just stand and marvel’ at it. If so, then ‘the brain must do whatever it takes to *assuage epistemic hunger* [. . .]. If our brains can just satisfy all our particular epistemic hungers as they arise, we will never find ground for complaint’.

Thus, it is one thing to misperceive some actual object as exemplifying a property that the object does not really instantiate (illusion). It is another thing to have a visual experience in which ‘something totally unreal is conjured up’ (delusion). Only in the latter case, would one fail to stand in some relation to a mind-independent object. Not only does the ‘argument from illusion’ trade on the confusion between visual illusions and visual hallucinations, but the conclusion of the argument presupposes that in all cases of visual experiences, veridical as well as non-veridical, some ‘object’ must exist. Since in non-veridical hallucinatory experiences, a mind-independent ‘material’ object fails to exist, according to the argument from illusion, it follows that some mind-dependent (purely mental) object must be present in non-veridical cases. Finally, since the visual appearances are allegedly indistinguishable whether the experience is veridical or not, the conclusion is that in all cases of visual perception, what one perceives is a mental sense-datum.

## 2.2 Disjunctivism and the rejection of an interface between mind and world

In virtue of what Putnam (1994: 445–6) calls ‘a familiar pattern of recoil that causes philosophy to leap from one frying pan to fire, from fire to a different frying pan, from a different frying pan to a different fire’, sense-datum theories have prompted a ‘direct’ or ‘naive’ realist response. Once sense-data are postulated as mental intermediaries between the human mind and mind-independent objects, it seems as if knowledge of mind-independent objects will forever remain inaccessible. Thus, the direct realist ‘recoil’ is a response to the threat of scepticism involved in sense-datum theories. As Martin (2001: 12) puts it, ‘a familiar objection to sense-datum theories of perception is that they introduce entities which act as a “veil of perception” between us and the external world; and it is often suggested that the putative presence of such a veil would lead to insuperable sceptical problems’. The threat is that, if all we are aware of in visual perception is the ‘veil’ of mental representations, then knowledge of mind-independent

<sup>10</sup>In some pathological cases, hallucinations can be stronger than Dennett seems to allow. See in Chapter 3 our discussion of somatoparaphrenia subsequent to right parietal lesions.

objects is bound to escape us. As Putnam (1994: 453) says on behalf of direct (or 'naive') realism, 'the disaster is the idea that there has to be an interface between our cognitive powers and the external world—or, to put the same point differently, the idea that our cognitive powers cannot reach all the way to the objects themselves'.

In order to avoid the threat of scepticism, direct realists espouse what they call a 'disjunctive' account of visual experience. The disjunctive account is so-called because it claims that visual experiences differ according to whether they are veridical or not. On this view, there is no common factor between veridical and non-veridical visual experiences. According to McDowell (1982), a leading exponent of 'disjunctivism', so-called 'highest common factor' conceptions of visual appearances are internalist theories: they rely on the alleged subjective indistinguishability between veridical and non-veridical cases of visual perception. What makes the sense-datum theory a 'highest common factor' conception of visual appearances is that, according to the sense-datum theory, there is a unique mental state that is the 'highest common factor' between veridical and non-veridical cases of visual perception. According to the sense-datum theory, there is a common 'narrow' phenomenological subjective content shared by veridical and non-veridical visual experiences, which consists in having in mind or perceiving a mind-dependent entity, i.e. a visual sense-datum.

It is quite clear that McDowell's motivation for espousing a disjunctive account of visual appearances is to circumvent the risk of scepticism involved in postulating a veil of mental intermediaries between the human mind and the world of mind-independent objects. According to disjunctivism, the 'highest common factor' conception of visual appearances leads to scepticism. Thus, it is the task of disjunctivism to deny the existence of mental representations with a 'narrow' content common to veridical and to non-veridical visual experiences in order to avoid the threat of scepticism. The disjunctive account makes a sharp distinction among visual experiences according to whether they make a mind-independent fact manifest to the mind or not. On the disjunctive account, there is a gap between the fact that an object looks a certain way to an observer and its seeming to an observer as if something looks a certain way. As McDowell (1982: 211) puts it:

[...] an appearance that such-and-such is the case can be *either* a mere appearance *or* the fact made manifest to someone [...] the object of experience in the deceptive case is a mere appearance. But we are not to accept that in the non-deceptive case too the object of experience is a mere appearance, and hence something that falls short of the fact itself. On the contrary, we are to insist that the appearance that is presented to one in those cases is a matter of the fact itself being disclosed to the experiencer. So appearances are no longer conceived as in general intervening between the experiencing subject and the world.

As Snowdon (1980–81: 186) puts it:

[...] the disjunctive picture divides what makes looks ascriptions true into two classes. In cases where there is no sighting they are made true by a state of affairs intrinsically independent of surrounding objects; but in cases of sightings the truth-conferring state of affairs involves the surrounding objects.

On this view then, there is a primitive contrast between the veridical perception of a mind-independent fact and cases of non-veridical perception where no mind-independent fact is made manifest to the human mind.

Thus, disjunctivism starts from the radical metaphysical realist assumption that the goal of visual perception is to make mind-independent facts available to the human mind. In order to capture the very idea of a mind-independent fact being made manifest to the human mind, McDowell (1982) appeals to the idea of the mind's 'direct openness to the world'. This idea of the human mind being directly open to the world is the precursor of Putnam's (1994) claim that 'the disaster is the idea that there has to be an interface between our cognitive powers and the external world'. Only if human minds are distinct from mind-independent facts can there be an interface between the former and the latter. Only at the cost of denying the interface between the world and the human mind can the latter be 'directly open' to the former. Indeed, in his more recent writings, McDowell (1994) seems to give up the very idea of mind-independent facts altogether: the world has become mind-dependent.

In his recent writings, McDowell (1994) rejects what he calls 'bald naturalism' and argues for what he calls a 're-enchanting' version of naturalism. According to McDowell (1994), 'bald naturalism' relies on the dualism between 'the space of natural law' and 'the space of human reasons'. This dualism has given rise to a picture in which nature turns out to be 'disenchanted'. Against the bald naturalistic dualism between the realm of law and the realm of reasons, McDowell (1994: 85–8) recommends explicitly that we replace the 'disenchanted' picture of nature by what he calls 'second nature', in which 'meaning is not a mysterious gift from outside nature'. Only by projecting the mind onto the world can the limitations of bald naturalism be overwhelmed. Only if the world is itself mind-dependent, it seems, can what Grice (1957) called 'non-natural meaning' be part of the world. But then if non-natural meaning is as much in the natural world as natural meaning is, why call the former *non-natural*? As we argued above, as natural meaning and natural signs arise from correlations, they are ubiquitous. Non-natural meaning or the content of representations is not. So it seems as if one would avoid the threat of scepticism at the cost of denying both the very notion of mind-independent objects and the very distinction between information and representation. McDowell (1994) does not offer a solution to the threat of scepticism consistent with metaphysical realism. He gives up metaphysical realism. If so, then the question is: how much comfort is it?

If one thinks, as we do, that 're-enchantment' of the kind McDowell (1994) is urging, is simply not compatible with naturalism, then one will better stick to the idea of an interface between mind and the world. We like RTVM precisely because we think that it avoids the pitfalls of both sense-datum theory and disjunctivism. RTVM is often called 'intentionalism' because it makes the basic claim that the content of a visual experience crucially depends upon the 'intentional' properties that the experience represents objects as having. According to the teleosemantic version of visual intentionalism sketched above, the human visual system has been selected by evolution for carrying information about (or for processing) a particular class of properties instantiated in the environment of early humans. According to visual intentionalism, what matters to a visual experience—what makes it the experience that it is—are the *properties* that are represented in the experience, not the particular objects that happen to exemplify the properties. What matters are the properties to which the visual system has been tuned by evolution. This is why in visual hallucinations, the visual mind can only 'conjure up' the representation of *visual* properties, not the representation of auditory, olfactory or tactile properties. In non-veridical visual experiences, the visual

mind can only represent misleadingly the properties that it is its function to carry information about.

Visual intentionalism and disjunctivism advocated by direct realists both reject the two basic claims of sense-datum theory, namely that in visual perception, the mind is acquainted with mind-dependent sense-data and that all knowledge of mind-independent facts is derivative from the mind's acquaintance with sense-data. Furthermore, visual intentionalism and direct realism reject an introspective implication of the sense-datum theory. According to the sense-datum theory, as alluded to in Russell's (1911) quotation above, introspection should make one aware of the subjective qualities of one's visual experiences (or sense-data). But as advocates of visual intentionalism, Harman (1990) and Tye (1992), have pointed out, visual experiences are 'transparent' or 'diaphanous'. As Martin (2001: 3) writes:

When I stare at the straggling lavender bush at the end of my street, I can attend to the variegated colors and shapes of leaves and branches [. . .]. But I can also reflect on what it is like for me now to be staring at the bush [. . .]. When my attention is directed out at the world, the lavender bush and its features occupy center stage. It is also notable that when my attention is turned inwards instead to my experience, the bush is not replaced by some other entity belonging to the inner realm of the mind in contrast to the run-down public sphere of North London.

Introspection of one's visual experience does not reveal any intrinsic property of one's experience: what it reveals is what the experience is *about* or what the experience is an experience *of*. So far, visual intentionalism and the disjunctivism of direct realism can agree against the sense-datum theory.

Thus, our brand of visual intentionalism accepts, while sense-datum theory denies, that mind-independent objects *can* be presented to the mind in visual experiences. According to visual intentionalism, unless such visual properties as shape, orientation, color, texture, motion had been instantiated by mind-independent objects in the environment of early humans, the human visual system could not have carried information about them, let alone been selected to do so by evolution. Our brand of visual intentionalism denies, while sense-datum theory asserts, that what is represented in visual experiences *must* actually exist. According to sense-datum theory, what exists in all cases is a mental (or mind-dependent) entity. Visual intentionalism accepts it that in visual delusion (or hallucination), nothing exemplifies the properties that are represented in the visual experience. Disjunctivism agrees with visual intentionalism, against sense-datum theory, that in visual experience, mind-independent facts and objects *can* be represented. According to disjunctivism, however, mind-independent facts and objects *must* be represented for a visual experience to be a genuine visual experience. On this point, visual intentionalism and disjunctivism part company: according to visual intentionalism, visual experiences are *not* required to make mind-independent facts manifest to the mind. They can, but they do not have to.<sup>11</sup>

<sup>11</sup>In Section 2.2, our goal has been to dissociate our version of the representational theory of the visual mind ('visual intentionalism') from both 'direct realism' and 'sense-datum' approaches. In particular, we have addressed the worry that 'visual intentionalism' might commit us to some version of the 'sense-datum fallacy'. We have not yet addressed directly the view of anti-representationalist philosophers who think that there are subjective, phenomenal non-representational properties of visual experiences or that its representational properties do not exhaust the phenomenal character of a visual experience. Nor do we claim that anti-representationalism in this sense is committed to the 'sense-datum fallacy'. We shall deal with a version of anti-representationalism, which we call phenomenal realism, at the end of Chapter 5.

### 2.3 The challenge of the homunculus

Finally, RTVM raises a puzzle: the puzzle of the homunculus. A non-mental representation (an artifact), such as a painting, has both representational and non-representational intrinsic physical and chemical properties. A physical picture will not reveal its non-mental pictorial content unless an intentional observer looks at it and his or her visual system causally interacts with the picture's intrinsic properties. As Ramachandran and Blakeslee (1998: 66) have noticed, the model of the physical picture leads to the assumption that 'there is a screen somewhere inside the brain where images are displayed'. This assumption in turn leads to an infinite regress for 'if you were to display an image [...] on an internal neural screen, you'd need another little person inside the brain to see that image. And that won't solve the problem either because you'd need yet another, even tinier person inside his head to view that image, and so on and so forth, ad infinitum'. So the puzzle for the representational theory of the visual mind is: how could anything in a person's visual brain be a mental representation unless a homunculus sitting in the person's brain looks at it and perceives its intrinsic neurophysiological properties? How could the discharge of neurons in a person's visual cortex represent anything unless something or someone could perceive their intrinsic properties?

The answer to this puzzle is that, according to RTVM, there is a basic asymmetry between mental representations and representational artifacts. Perceiving a representational artifact consists in detecting its intrinsic properties: one will not be aware of the content of a representational artifact unless one perceives its intrinsic properties. To be visually aware of the content of a picture is to come to know a fact about the picture by visual means. To be visually aware of the content of a picture is to know (hence believe) that it represents, e.g. a ship, a tiger, a house or a tree. In Chapter 5, we shall call this knowledge 'primary epistemic seeing'. One comes to know what a picture represents by seeing the picture: i.e. in virtue of non-epistemically seeing shapes, contours and patches of colors on a two-dimensional canvas. The visual experience consists in seeing the two-dimensional shapes, contours and patches of colors. One cannot see shapes, contours and patches of colors unless one is caused to do so in virtue of responding to the intrinsic properties of the canvas.

Like representational artifacts, mental representations too have intrinsic properties. However—and this is crucial—contrary to the sense-datum theory, according to visual intentionalism, when one visually perceives two-dimensional colors and shapes laid out on a canvas, there is literally nothing *in* the perceiver's brain with the properties of the canvas: there is no sense-datum in the perceiver's brain with the colors and shapes of the canvas. In the brain, there is only electrical activity. So in processing a visual stimulus, one becomes aware of properties of the canvas and the processing in the brain is patterns of electrical activity—it is not an inner perception of mental sense-data. In processing a visual stimulus, one does not become aware of one's percept. The percept represents a physical object and the formation of the percept is the result of the electrical activity of the brain.

This is not to say that one cannot become aware of one's mental representations. For one thing, with the use of sophisticated contemporary techniques of brain imagery, it becomes possible to detect the intrinsic properties of mental representations and as a result to visualize the activities of the brain. But visualizing the activity of a brain area



during a perceptual task is not visualizing what is perceived in the task. For example, if a subject perceives a visual token of the word 'cat' in a task of brain imagery, by perceiving an image of the subject's brain activity during the task, one will not thereby see that what the subject was perceiving was a token of the word 'cat'. Nor, as Harman (1990) and Tye (1992) have emphasized, does visual perception consist in becoming aware of one's own visual percept: in Harman's (1990) words, it is a fallacy (the 'sense-datum fallacy') to assume that 'one's awareness of the color of a strawberry is mediated by one's awareness of an intrinsic feature of a perceptual representation'. In perceiving a lavender bush, one is aware of the bush, not of the percept. This is what we called above the 'transparency of experience': visual percepts are 'diaphanous'.<sup>12</sup>

The basic asymmetry between mental representations and representational artifacts can be stated thus: one cannot be aware of the pictorial content of a picture unless one detects the intrinsic properties of the picture. One becomes visually aware of a tree by forming a visual percept of a tree. This process would not be possible unless one were detecting intrinsic properties of the tree. As a result, one visually perceives properties of the tree, not properties of one's visual percept. To say, as we did above, that representational artifacts derive their contents from the primitive contents of mental representations, is to accept the intentionalist thesis that the perceptual process, whereby one becomes aware of the pictorial content of a picture, consists in the formation and transformation of mental representations. In this process, however, mental representations themselves are not perceived visually or otherwise by anyone or anything. In ordinary visual perception, one does *not* become aware of one's own mental representations: the formation and transformation of mental representations is the process whereby one becomes aware of things and properties in one's environment. Intentionalists can agree with Putnam (1994: 453) when he says: 'We don't *perceive* visual experiences, we *have* them'.

According to visual intentionalism, one becomes aware of properties exemplified in one's environment by means of the mental perceptual representations supplied by one's visual system. In visual perception, one becomes visually aware of one's environment through the process whereby mental representations are formed and transformed. In this process, one does not become aware of them. This is not to deny that by introspection one may become aware of the contents of one's own mental representations. But if one can and if one does, then, according to the representational theory of the mind, one is not made aware of one's mental representations by perceiving them (visually or otherwise). One does not visually perceive one's visual percepts any more than one hears one's auditory percepts. One can only form higher order beliefs about them. One can only come to believe that one is enjoying such and such a visual experience. In having a visual experience, one sees objects, properties and facts (many of which are mind-independent). One's visual experience itself is not something to be seen—neither by oneself nor by anybody else (unless one uses brain imaging techniques). As argued by Rosenthal (1986, 1993) and Shoemaker (1994), one's introspective knowledge of one's own visual mind is metarepresentational, not perceptual.

<sup>12</sup>The claim of transparency is restricted to visual perception. We are not claiming that the human mind is introspectively blind to the differences between perceiving, believing, remembering, desiring and intending.

### 3 Conceptual and non-conceptual content

Mental representations come in several varieties. As we said above, thoughts, judgments, beliefs, desires, perceptual experiences and mental images all are mental representations. Much recent philosophy of mind and perception has been devoted to the distinction between the conceptual contents of thoughts and the non-conceptual contents of perceptual representations. Although the issue is somewhat controversial, we accept the distinction between conceptual and non-conceptual content.<sup>13</sup>

The paradigmatic arguments in favor of the assumption that perceptual representations possess a distinctive kind of content advert to three basic kinds of considerations: first of all, they emphasize the fine-grainedness and the richness of the phenomenology of visually formed perceptual representations, which cannot be matched by concepts mastered by the person having the visual experience. Second, they reject the requirement that unless a creature possesses appropriate concepts matching the detailed texture of her perceptual experience, she will not enjoy the perceptual experience. Finally, they point towards a distinctive link between the fine-grainedness of the non-conceptual content of perceptual experience and the fine-tuning of object-oriented actions. We accept both the distinction between the conceptual content of thought and the non-conceptual content of perceptual experience, and the first pair of considerations in its favor. We subscribe, however, to the fundamental duality between the visual perception of objects and visually guided actions directed towards objects. Therefore, as we already argued in the Introduction, we do not accept what Clark (2001) dubs ‘the assumption of experience-based control’ (EBC), i.e. the idea of a constitutive link between the non-conceptual content of visual perceptual experiences and the fine-tuning of visually guided actions.<sup>14</sup>

#### 3.1 The productivity and systematicity of thoughts

As many philosophers of mind and language have argued, what is characteristic of conceptual representations is that they are both productive and systematic. Like sentences of natural languages, thoughts are productive in the sense that they form an open-ended infinite set. Although the lexicon of a natural language is made up of finitely many words, thanks to its syntactic rules, a language contains indefinitely many well-formed sentences. Similarly, an individual may entertain indefinitely many conceptual thoughts. In particular, both sentences of public languages and conceptual thoughts contain such devices as negation, conjunction, disjunction and conditionals. So one can form indefinitely many new thoughts by prefixing a thought by a negation operator, by forming a disjunctive, a conjunctive or a conditional thought out of two simpler thoughts or one can generalize a singular thought by means of quantifiers. Sentences of natural languages are systematic in the sense that if a language contains a sentence *S* with a syntactic structure, e.g. *Rab*, then it must contain a sentence *S'* expressing a syntactically related sentence, e.g. *Rba*. An individual’s conceptual thoughts are supposed

<sup>13</sup>For opposition to the distinction, see e.g. McDowell (1994).

<sup>14</sup>An idea that Clark (2001) rejects.

to be systematic too: if a person has the ability to entertain the thought that, for example, John loves Mary, then she must have the ability to entertain the thought that Mary loves John. If a person can form the thought that *Fa*, then she can form both the thought that *Fb* and the thought that *Ga* (where ‘*a*’ and ‘*b*’ stand for individuals and ‘*F*’ and ‘*G*’ stand for properties). Both Fodor’s (1975, 1987) language of thought hypothesis and Evans’ (1982) generality constraint are designed to account for the productivity and the systematicity of thoughts, i.e. conceptual representations. It is constitutive of thoughts that they are structured and that they involve conceptual constituents that can be combined and recombined to generate indefinitely many new structured thoughts. Thus, concepts are building blocks with inferential roles.

Because they are productive and systematic, conceptual thoughts can rise above the limitations imposed to perceptual representations by the constraints inherent to perception. Unlike thought, visual perception requires some causal interaction between a source of information and some sensory organs. For example, by combining the concepts *horse* and *horn*, one may form the complex concept *unicorn*, even though no unicorn has or ever will be visually perceived (except in visual works of art). Although no unicorn has ever been perceived, within a fictional context, on the basis of the inferential role of its constituents, one can draw the inference that if something is a unicorn, then it has four legs, it eats grass and it is a mammal.

Hence, possessing concepts must involve the mastery of inferential relations: the latter is an important part of the former.<sup>15</sup> Only a creature with conceptual abilities can draw consequences from her perceptual processing of a visual stimulus. Following Dretske (1969), what we shall call ‘epistemic seeing’ in Chapter 5, requires conceptual processing of perceptual inputs. Thought and visual perception are clearly different cognitive processes. One can think about numbers and one can form negative, disjunctive, conjunctive and general thoughts involving multiple quantifiers. Although one can visually perceive numerals, one cannot visually perceive numbers. Nor can one visually perceive negative, disjunctive, conjunctive or general facts (corresponding to e.g. universally quantified thoughts).

As Crane (1992a: 152) puts it, ‘there is no such thing as deductive inference *between* perceptions’. Upon seeing a brown dog, one can see at once that the animal one faces is a dog and that it is brown. If one perceives a brown animal and one is told that it is a dog, then one can certainly come to believe that the brown animal is a dog or that the dog is brown. But on this hybrid epistemic basis, one thinks or believes, but one does not *see* that the dog is brown. One came to know that the dog is brown by seeing it. But one did not come to know that what is brown is a dog by seeing it.<sup>16</sup> Unlike the content of concepts, the content of visual percepts is not a matter of inferential role. As emphasized by Crane (*ibid.*), this is not to say that the content of visual percepts is amorphous or unstructured. One proposal for capturing the non-conceptual structure of visual percepts is Peacocke’s (1992a) notion of a scenario content, i.e. a visual way of filling in space. As we shall see momentarily, one can think or believe of an animal that it is a dog without thinking or believing that it has a color. But one

<sup>15</sup>Not everyone agrees. Fodor (1998b), e.g. vehemently denies it and espouses an atomic theory of the contents of concepts.

<sup>16</sup>For extended discussion of the scope and limits of *visual* knowledge, see Chapter 5.

cannot see a dog in broad daylight without seeing its color. We shall try to capture this feature of the content of visual percepts, which is part of their distinctive informational richness, in terms of the distinctive function of visual perceptual mechanisms to produce states that encode information analogically.

### 3.2 The fine-grainedness and informational richness of visual percepts

Unlike thought, visual perception has a spatial, perspectival, iconic and/or pictorial structure not shared by conceptual thought. Arguably, one can visually perceive dots with no spatial internal structure, but one cannot visually perceive an object unless it has some spatial location or other. The content of visual perception has a spatial perspectival structure that pure thoughts lack. In order to apply the concept of a dog, one does not have to occupy a particular spatial perspective relative to any dog. But one cannot see a dog unless one occupies some spatial standpoint or other relative to it: one cannot see a dog simultaneously from the top and from below, from the front and from the back. The concept of a dog applies indiscriminately to poodles, alsatians, dalmatians or bulldogs. One can think that all European dogs bark. But one cannot see all European dogs bark, let alone all dogs in the world bark. Nor can one see a generic dog bark. One must see some particular dog: a poodle, an alsatian, a dalmatian or a bulldog, as it might be. Although one and the same concept—the concept of a dog—may apply to a poodle, an alsatian, a dalmatian or a bulldog, seeing one of them is a very different visual experience than seeing another. One can think that a dog barks without thinking of any other properties of the dog. One cannot, however, see a dog in broad daylight unless one sees its shape and the colors and texture of its hairs. Of course, in poor illumination conditions (e.g. at night), one might see the shape of a dog while failing to see its color. In still worse illumination conditions, one may also become unable to see the shape and spatial orientation of a dog.

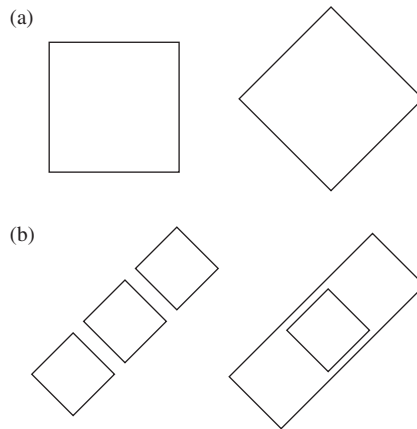
Thus, the content of visual perceptual representations turns out to be both more fine-grained and informationally richer than the conceptual contents of thoughts. There are three paradigmatic cases in which the need to distinguish between conceptual content and the non-conceptual content of visual perceptions may arise. First, a creature may be perceptually sensitive to objective perceptual differences for which she has no concepts. There may be something it is like to enjoy a visual experience of a shape and/or a color for which the creature has no concept. Second, two creatures may enjoy one and the same visual experience, which they may be inclined to conceptualize differently. Finally, two different persons may enjoy two distinct visual experiences in the presence of one and the same distal stimulus to which they may be inclined to apply one and the same concept.

Peacocke (1992b: 67–8) asks us to consider, e.g. a person's visual experience of a range of mountains. As he notices, one might want to conceptualize one's visual experience with the help of concepts of shapes expressible in English with such predicates as 'round' and 'jagged'. But these concepts of shapes could apply to the non-conceptual contents of several different visual experiences prompted by the distinct shapes of several distinct mountains. Arguably, although a human being might not possess any concept of shape whose fine-grainedness could match that of her visual experience of the shape of the mountain, her visual experience of the shape is

nonetheless distinctive and it may differ from the visual experience of the distinct shape of a different mountain to which she would apply the very same concept. Similarly, human beings are perceptually sensitive to far more colors than they have color concepts and color names to apply. Although a human being might lack two distinct concepts for two distinct shades of color, she might well enjoy a visual experience of one shade that is distinct from her visual experience of the other shade. As Raffman (1995: 295) puts it, ‘discriminations along perceptual dimensions surpasses identification [...] our ability to judge whether two or more stimuli are the same or different surpasses our ability to type-identify them’.

Two persons might enjoy one and the same kind of visual experience prompted by one and the same shape or one and the same color, to which they would be inclined to apply pairs of distinct concepts, such as ‘red’ vs. ‘crimson’ or ‘polygon’ vs. ‘square’. If so, it would be justified to distinguish the non-conceptual content of their common visual experience from the different concepts that each would be willing to apply. Conversely, as argued by Peacocke (1998), presented with one and the same geometrical object, two persons might be inclined to apply one and the same generic shape concept, e.g. ‘that polygon’, and still enjoy different perceptual experiences or see the same object as having different shapes. For example, as Peacocke (1998: 381) points out, ‘one and the same shape may be perceived as square, or as diamond-shaped [...] the difference between these ways is a matter of which symmetries of the shape are perceived; though of course the subject himself does not need to know that this is the nature of the difference’. If one mentally partitions a square by bisecting its right angles, one sees it as a diamond. If one mentally partitions it by bisecting its sides, one sees it as a square. Presumably, one does not need the concept of an axis of symmetry to perform mentally these two bisections and enjoy two distinct visual experiences. Arguably, one and the same object with one and the same shape can be perceived as a diamond or as a square. Arguably, seeing it as a square or as a diamond may depend on its orientation relative to the perceiver. So a square tilted at  $45^\circ$  will look like a diamond. But as Peacocke (1992b: 76) points out, if embedded within a larger rectangle sharing the same orientation, the very same object will look like a square tilted at  $45^\circ$  (Fig. 1.1).

Now, the rotation of an object around an axis preserves its shape. So if changing the orientation of an object may change the content of the visual perception of the object, then presumably the shape property itself is not sufficiently fine-grained to individuate the non-conceptual content of the visual experience of the object. In addition to the shape property, we need what Peacocke (1989) calls a particular ‘perceptual mode of presentation’ of the shape. We earlier mentioned the fact that the way one and the same shape is presented by visual perception differs from the way it is presented by, e.g. touch. For instance, in a visual perceptual experience, the representation of shape interacts with the representation of colors, texture and orientation in a way that is unique to the visual modality. This special kind of interaction among the representations of specific properties, which is unique to normal visual perception, is part of the visual perceptual mode of presentation of the shape property. Arguably, being a square and being a diamond are two ways of exemplifying one and the same shape property, i.e. the same way of being shaped. In Peacocke’s (2001) terms then, ‘a way in which a shape property may be perceived is to be [...] distinguished from a way of being shaped’. Again and again in Chapter 4, we shall see instances of visual processing of illusory displays



**Fig. 1.1** One and the same objective shape can prompt different visual perceptual experiences: (a) the same shape in different orientations is perceived as a square (*left*) or as a regular diamond (*right*); (b) in the proper context, what was perceived as a regular diamond in (a) can be perceived as a tilted square. (From Peacocke 1992a permission sought.)

such that the visual mode of presentation of one and the same shape property (e.g. of a Titchener central disk surrounded by an annulus of either smaller or larger circles than it) varies according to whether the task is a perceptual task or a motor task.

Against this kind of argument in favor of the non-conceptual content of visual experiences, McDowell (1994, 1998) has argued that demonstrative concepts expressible by, e.g. ‘that shade of color’ or ‘that shape’, are perfectly suited to capture the fine-grainedness of the visual percept of color and/or shape. We are willing to concede to McDowell that such demonstrative concepts do exist and play an important role in our perceptually based thoughts.

But first of all, we agree with Peacocke (2001) that, unlike perceptual modes of presentation of properties, demonstrative concepts are by their very nature true of the properties of particular objects. In thinking about the shape of an object by means of such a demonstrative concept as ‘that shape’, one is latching onto a particular object that is demonstrated (or demonstratively referred to) and that happens to exemplify the property expressed by the concept ‘shape’. In such a demonstrative thought, the object referred to is being visually perceived, but the shape property is being *thought* about. In fact, the object is not merely being perceived, it is also being *referred* to. The property is being thought about while the object is being visually perceived and referred to. For the concept ‘that shape’ to be a demonstrative concept, the general concept expressed by ‘shape’ must apply to the property of a particular object that is both perceived and referred to. Although one cannot perceive the shape of an object unless one perceives the object exemplifying the shape property, what is constitutive of the visual perceptual mode of presentation of the shape of an object is that one represents the shape pictorially. In a visual perceptual mode of presentation, the shape is not thought about; it is visually experienced and pictorially encoded with other visual attributes of the object

such as its size, orientation and color. Nor does one have to refer to an object in order to visually experience its shape. Arguably, one cannot visually experience the shape of an object unless one perceives the object. But perceiving an object is one thing; referring to it is another thing. In perceptual experience, information flows from the object towards the perceiver: the object represented causes the representation. Although thought is not always an intentional action, reference in general and demonstrative reference in particular are intentional actions. One's intention to refer causes an object to be referred to. Even attending to an object via visual perception is not the same thing as referring to it. Attending to an object is part of perception. Referring to an object is part of thought.

Second, we agree with Bermudez (1998: 55–7) and Dokic and Pacherie (2001) that such demonstrative concepts would seem to be too weak to perform one of the fundamental jobs that color concepts and shape concepts must be able to perform—namely recognition. Color concepts and shape concepts stored in a creature's memory must allow recognition and re-identification of colors and shapes over long periods of time. Although pure demonstrative color concepts may allow comparison of simultaneously presented samples of color, it is unlikely that they can be used to reliably reidentify one and the same sample over time. Nor presumably could pairs of demonstrative color concepts be used to reliably discriminate pairs of color samples over time. As we shall argue in Section 4, in a perceptual episode, just as one can track the spatio-temporal evolution of a perceived object, one can store in a temporary object file information about its visual properties in a purely indexical or demonstrative format. If, however, information about an object's visual properties is to be stored in episodic memory, for future re-identification, then it cannot be stored in a purely demonstrative or indexical format, which is linked to a particular perceptual context. At a minimum, the demonstrative must be fleshed with descriptive content. This is part of what Raffman (1995: 297) calls 'the memory constraint'. As Raffman (1995: 296) puts it:

[...] the coarse grained character of perceptual memory explains why we can recognize 'determinable' colors like red and blue and even scarlet and indigo *as such*, but not 'determinate' shades of those determinables [...]. Because we cannot recognize determinate shades as such, *ostension* is our only means of communicating our knowledge of them. If I want to convey to you the precise shade of an object I see, I must point to it, or perhaps paint you a picture of it [...]. I must present you with an instance of that shade. You must have the experience yourself.

Now, if the conceptualist was tempted to turn the tables around and argue that demonstrative concepts (of shapes or colors) are precisely well-suited to capture the fine-grainedness of perceptual experiences on the grounds that they are not designed to achieve recognitional tasks, we would really ask in what sense they would still deserve to be called concepts. If the link between the mental pointer of a shape or a color property and memory is so weak as to preclude re-identification of the shape or the color property, the mental pointer hardly counts as a concept.

The distinctive informational richness of the content of visual percepts has been discussed by Dretske (1981) in terms of what he calls the analogical coding of information.<sup>17</sup>

<sup>17</sup>For discussion, see Jacob (1997, ch. 2).

One and the same piece of information—one and the same fact—may be coded analogically or digitally. In Dretske's sense, a signal carries the information that, e.g.  $a$  is  $F$  in a digital form if the signal carries no additional information about  $a$  that is not already nested in the fact that  $a$  is  $F$ . If the signal does carry additional information about  $a$  that is not nested in the fact that  $a$  is  $F$ , then the information that  $a$  is  $F$  is carried by the signal in an analogical (or analog) form. For example, the information that a designated cup contains coffee may be carried in a digital form by the utterance of the English sentence: 'There is some coffee in the cup'. The same information can also be carried in an analog form by a picture or by a photograph. Unlike the utterance of the sentence, the picture *cannot* carry the information that the cup contains coffee without carrying additional information about the shape, size, orientation of the cup and the color and the amount of coffee in it. As we pointed out above, unlike the concept of a dog, the visual percept of a dog carries information about which dog one sees, its spatial position, the color and texture of its hairs, and so on.

There are at least two important and related reasons why Dretske's (1981) distinction between the analog and the digital encoding of one and the same piece of information seems unsuitable as a basis for capturing the distinction between the informational richness of the non-conceptual content of perceptual experience and the conceptual content of thought. First of all, in Dretske's (1981) view, what can be encoded either analogically or digitally is *informational* content, *not representational* content. Earlier in this chapter, we expressed our agreement with a teleological view of representational content according to which, to represent the fact that  $x$  is  $F$  is to have the *function* to indicate (or carry information about) the presence of  $F$ s. Why is this important? Because unlike mere informational content, representational content has *correctness* conditions: unlike a signal carrying information, a representation can be correct or incorrect. If the latter, it is a misrepresentation. As argued forcefully by Millikan (1984, 1993, 2000) and recognized by Dretske (1988, 1995b), unless a system has a function, it cannot misrepresent anything. If so, then Dretske's (1981) analog/digital distinction seems to apply to the wrong kind of thing, namely informational content, not representational content with correctness conditions.

Second, it seems as if any representational state whatsoever can be said to encode some information analogically and some information digitally. For instance, the thought that  $x$  is a square can be said to represent conceptually and encode digitally the fact that  $x$  is a square and it can be said to encode analogically the fact that  $x$  is a rectangle.<sup>18</sup> Presumably, a visual experience can be a non-conceptual representation of, and encode digitally, the fact that, e.g.  $x$  presents a certain spatial visual appearance.<sup>19</sup> The same visual experience could also be said to encode analogically the fact that  $x$  is a rectangle. Thus, the thought that  $x$  is a square and a visual experience can encode analogically the same piece of information, namely the fact that  $x$  is a rectangle. If so, then the analog/digital distinction in Dretske's (1981) sense seems to cut across the conceptual/non-conceptual content distinction.

<sup>18</sup>As noticed by Peacocke (1989: 315), Millikan (2000) and Pacherie (2000a).

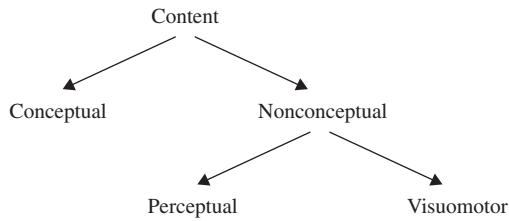
<sup>19</sup>See Peacocke's (1992b) notion of a scenario content.



Perhaps one could nonetheless try to reconcile Dretske's (1981) analog/digital distinction together with a teleological approach to the distinction between the conceptual/non-conceptual content distinction along the following lines. Consider the fact that  $x$  is a square. Whereas the thought that  $x$  is a square is a state produced by a mechanism with the function to carry the information that  $x$  is a square in digital form, one's visual experience prompted by perceiving the very same fact (or state of affairs) would be a state produced by a mechanism with the function to carry information about it (the very same state of affairs) in analogical form. The analog/digital distinction would then serve to capture some of the difference between the non-conceptual content of a visual experience and the conceptual content of a thought within a teleological framework. But the teleological distinction between mechanisms with a function to carry information in analog and in digital form, respectively, would itself be relativized to particular facts (or states of affairs). As we shall argue in Chapters 5 and 6, the teleological distinction between states produced by mechanisms with the function to carry information in analog and in digital form, respectively, is consistent with what we shall soon call the constraint of *contrastive identification*.

To sum up, the arguments by philosophers of mind and by perceptual psychologists in favor of the distinction between the conceptual content of thought and the non-conceptual content of visual percepts, turn on two basic considerations: on the one hand, they rely on the rejection of the claim that unless a creature has the conceptual resources appropriate to conceptualize her experience, she will not enjoy a perceptual visual experience at all; on the other hand, they rely on the distinctive fine-grainedness and the informational richness of visual percepts. Thus, the second considerations turn on the phenomenology of visual experience. In Chapters 2, 3 and 4, we shall provide a variety of evidence from electrophysiological and behavioral studies on macaque monkeys, the neuropsychological examination of brain-lesioned human patients and from psychophysical experiments performed on normal human subjects that point to a different kind of non-conceptual content, which we shall label 'visuomotor' content. In Chapter 6, we shall emphasize the fact that, unlike the arguments in favor of the non-conceptual content of visual percepts, the arguments for the distinction between the non-conceptual content of visual percepts and the non-conceptual content of visuomotor representations do not rely on phenomenology at all. Rather, they rely on the need to postulate mental representations with visuomotor content in order to provide a causal explanation of visually guided actions towards objects. Thus, we submit that the non-conceptual content of visual representations ought to be bifurcated into perceptual and visuomotor content as in Fig. 1.2.

In accord with what Clark's (2001) calls 'the assumption of experience-based selection' (EBS), we emphasize the link between the non-conceptual content of conscious visual experience, conceptual thought and memory: the non-conceptual content of visual experiences is involved in selecting a target for action. But we de-emphasize the connection between the non-conceptual content of visual experiences and visually guided actions towards objects, encapsulated by Clark's (2001) EBC assumption. By so doing, we do not thereby endorse the conceptualist view according to which the content of visual experiences is itself conceptualized. Unlike the content of visuomotor representations, the content of conscious visual experiences is poised for conceptual use, but it is not conceptual content.



**Fig. 1.2** Bifurcation of the nonconceptual content of visual representations into perceptual and visuomotor content.

In the next section, we shall lay the groundwork for Chapters 5 and 6 by examining a process we shall call ‘cognitive dynamics’. Cognitive dynamics is a two-fold process: on the one hand, it involves the mapping of visual percepts of objects and locations onto conceptual thoughts. On the other hand, it consists of cognitive adjustments from more engaged to more disengaged thoughts, and vice-versa.

## 4 Elements of cognitive dynamics

The expression ‘cognitive dynamics’ was first used by Kaplan (1989) in an influential piece of work (written and widely circulated several years earlier) devoted to the study of such context-dependent linguistic expressions as indexicals (e.g. ‘I’, ‘you’, ‘now’, ‘today’ ‘here’, ‘there’) and demonstratives (e.g. ‘he’, ‘this’, ‘that’, ‘there’). Indexicals and demonstratives are linguistic devices whose references change with the context of use. In Campbell’s (1994: 42–3) terminology, ‘I’ and ‘you’ are personal indexicals; ‘now’ and ‘today’ are temporal indexicals; ‘here’ and ‘there’ are spatial indexicals. Kaplan was interested in the linguistic adjustments one has to make in order to express one and the same thought at different times or in different places. Indeed, the problem had been raised in the early part of the twentieth century in a famous essay by the logician Gottlob Frege (1918–19) with respect to temporal thoughts: ‘If someone wants to say the same today as he expressed yesterday using the word “today”, he must replace this word with “yesterday”’.<sup>20</sup>

Evans (1982: 143–76) made an important contribution to the topic of cognitive dynamics by addressing the fundamental issue of what he called ‘demonstrative identification’, namely the human ability to track objects in one’s vicinity through perception and thought. As Evans (1982) emphasized, we may want to distinguish two major cases. During a single visual perceptual episode, one may track the trajectory of an object. Visual perception allows one to create or open a temporary object-file in which one

<sup>20</sup>See Dokic (1997) and Recanati (1997) for further discussion of what cognitive dynamics is about.

stores visual pictorial information about the object (e.g. its shape, color, texture). If and when the visual perceptual relation to the object is momentarily interrupted, the human episodic memory system allows one to recognize and re-identify an object over periods of time of variable length. Arguably, in the perceptual episode, the information gained by visual perception is directly linked to indexical and demonstrative concepts that can be used to refer to the perceived object from within an egocentric perspective. Arguably, in order for information gained by visual perception to be stored in episodic memory for future re-identification of an object, the information about the object must be stored in a more abstract, more descriptive format. One can refer to a perceptible object as ‘that sofa’ or even as ‘that’ (followed by no sortal), but presumably when one does not stand in a perceptual relation to the object, information about it cannot be stored in episodic memory in such a pure demonstrative format. Rather, it must be stored using a more descriptive symbol such as ‘the (or that) red sofa that used to face the fire-place’.

According to Evans (1982), a ‘past-tense’ demonstrative concept can be part of what he calls an ‘information-based’ thought as much as a perceptually-based demonstrative concept. On his view, it seems as if perception, episodic memory and testimony are all on a par at the service of the ‘informational system’. It seems to us, however, that there are important differences between the format in which information can be encoded in visual perception and in memory, respectively, as a result of the functional roles of perception and memory. Through perception one gains new information. Through memory, information is retained or preserved, not gained. Arguably, memory can modify the structure of information gained by perception: in memory, perceived items can be recombined so that what one remembers may differ from what one perceived. Memory can contribute to recreate one’s representation of one’s experienced past; but, if so, what one remembers is not true information. Thus, unlike perception, remembering cannot be a way of acquiring new knowledge.

As Evans emphasizes, one and the same English demonstrative pronoun ‘that’ can express a perceptually based demonstrative concept or a past-tense demonstrative concept. But on the one hand, language might be slightly misleading here: one and the same word might be used to express different concepts. We might use one and the same English demonstrative ‘that’ to refer to one and the same object to express slightly different concepts with relevantly different contents. To use an example of Evans (1982: 308), a perceptually based use of ‘that bird’ has a primarily spatial conceptual content used to refer to a bird accessible to perception. The perceptually based use of ‘that bird’ serves to direct attention to the spatial location of the perceptually accessible bird. The task of the audience of a perceptually based use of ‘that bird’ consists in a visual exploration of space to locate the perceptually accessible bird. A ‘past-tense’ use of the same demonstrative ‘that bird’ to refer to the same bird years later, as the bird is no longer accessible to perception, has a primarily temporal content. It serves to direct attention to a memory of the bird backwards in time. The task of the audience of a ‘past-tense’ use of ‘that bird’ is to search episodes stored in their episodic memory. Arguably, unlike the visual exploration of space, the mental exploration of one’s episodic memory, and the retrieval of a particular episode, require use of some descriptive information.

On the other hand, the claim about the difference between the format of the information gained by visual perception and the information stored in episodic memory (for future recall) applies primarily to information about objects (e.g. a bird or a sofa). We are claiming that the content of a demonstrative concept referring to an object supported by a perceptual link to that object must presumably be descriptively enriched in order for the information to be stored in episodic memory for future re-identification of the object referred to by the demonstrative concept. It may be, however, that the use of a ‘past-tense’ demonstrative (such as ‘that dinner party’ in the context of a memory retrieval) might serve to express a pure memory demonstrative concept available to one’s episodic memory, while one would be at a loss to provide further descriptive conceptual information for uniquely identifying which dinner party one is recalling. If so, we would claim that such pure memory demonstratives made available by the episodic memory experience itself apply to events, not to physical objects. In our view, there is a significant difference between perceiving/remembering a physical object and perceiving/remembering an event. The perceptual link to an object depends on a clear ontological demarcation between the object and the perceiver. By contrast, the perception of an event involves a different mereological relation between the perceiver (or the witness) and the event: the former is part of the latter. So in using a ‘past-tense’ demonstrative concept for an event stored in episodic memory, one is partly referring to oneself in a way one is not when one uses a ‘past-tense’ demonstrative concept of an object. This difference, we claim, accounts for why, unlike ‘pure’ memory demonstratives for events, memory demonstratives for physical objects must be fleshed with more descriptive information.<sup>21</sup>

Perry (1986a) made a useful distinction between vertical and lateral interpretations of sentences containing expressions whose meanings are more or less sensitive to the context. Interpreting a sentence *up* consists in rising above the context by replacing an expression whose meaning is more sensitive to the context by a coreferential expression whose meaning is less sensitive to the context. A person located in Lyon receives a post-card from a friend in San Francisco with the sentence: ‘This city has dilapidated cable-cars’. If the person in Lyon writes down in her notebook ‘San Francisco has dilapidated cable-cars’, she is interpreting *up*. Conversely, consider a person visiting San Francisco who reads the latter sentence in her travel-guide book. If she then writes down the former sentence in her diary, she is interpreting *down*. In other words, she replaces an expression whose meaning is less sensitive to the context by a coreferential expression whose meaning is more sensitive to the context. Lateral interpretation involves replacing an expression by a coreferential expression whose meaning is equally sensitive to the context (e.g. replacing ‘you’ by ‘I’).<sup>22</sup>

#### 4.1 Cognitive engagement and the detachment constraint on thoughts

In an interesting discussion of various ways of thinking about space, Campbell (1994: 5–6) distinguishes two broad ways of thinking of one and the same region of space: one

<sup>21</sup>Notice that one is not likely to refer to an event, e.g. a dinner party, of which one is both a witness and a constituent, by using the demonstrative ‘that’ with a spatial content.

<sup>22</sup>See Recanati (1997) for further elaboration of Perry’s distinctions.

way is ‘as a participant, as someone plunged into its center [...] with things to do in that space’. The other way is thinking about the same region of space ‘as a disengaged theorist’. The former engaged way of thinking about space is the way of thinking ‘one uses when sitting at a dinner table, moving and acting in that space’. The latter more detached way of thinking about the same region of space is the way of thinking ‘used subsequently by the detective who tries to reconstruct the scene and to establish who did what’. The distinction is that between ‘thinking of the space from a particular point of view, as a subject at the center of one’s world, and thinking about the space independently of any particular view point on it, in an impersonal or absolute way’.

In Section 3 we argued that thoughts satisfy Evans’ (1982) generality constraint and/or Fodor’s (1975, 1987) language of thought hypothesis. Thoughts involve concepts. Concepts have inferential roles and they can be recombined in indefinitely many new ways to form new thoughts. Unlike thoughts, visual percepts have non-conceptual iconic or pictorial contents. The pictorial non-conceptual content of visual percepts is informationally richer and more fine-grained than the conceptual contents of thoughts. Thus, the mapping from a visual percept onto a thought involves stripping the percept of much of its informational richness.

To see what is at issue in the mapping between a visual percept and a thought, consider an example adapted from Barwise (1989: 237). Let a visual percept represent a glass to the left of a bottle. The question is: ‘How can this visual percept give rise to a thought?’.<sup>23</sup> Notice that the mapping from visual percept to thought is one-way: visual percepts cause thoughts. Thoughts can cause other thoughts, but thoughts do not cause visual percepts.

As we argued above, unlike a thought, in a normal human subject, the visual percept cannot depict the location of the glass relative to the bottle without depicting the orientation, shape, texture, size and content (if any) of both the glass and the bottle. For the mapping to occur, however, the iconic or pictorial content of the visual percept representing the location of the glass relative to the bottle must match the conceptual content expressible by the English two place relation ‘being to the left of’. Conceptual processing of the pictorial content of the visual percept may yield a representation whose conceptual content can be expressed by the English sentence: ‘The glass is to the left of the bottle’. As we argued above, the conceptual content conveyed by an utterance is informationally impoverished relative to the informational richness of the pictorial content of the visual percept.

Now, once the visual percept has been turned into a thought by a process involving a selective elimination of information, further conceptual processing can yield a still more complex thought involving, not a two-place relation between pairs of objects, but a *three*-place relation between a pair of objects and an egocentric perspective. Once one has formed the thought that the glass is to the left of the bottle, it is a short step to form the more detached thought that the glass is to the left of the bottle from some egocentric perspective, not others. It is a short step to form the thought that, e.g. the glass is to the left of the bottle for someone looking at the fire-place but not for

<sup>23</sup>We shall pick up again this question from a slightly different angle in Chapter 5, Section 3, when we examine the scope and limits of visual knowledge and what Dretske calls ‘primary epistemic seeing’.

someone with his back onto the fire-place. From the latter perspective, the glass is to the right, not to the left, of the bottle. Here, we reach a fundamental difference between visual percepts and thoughts: one can move upwards from the thought involving the binary relation ‘being to the left of’, to the more detached thought involving the ternary relation ‘being to the left of from one’s egocentric perspective’. One can see a glass as being to the left of a bottle. But one cannot see a glass as being to the left of a bottle from one’s egocentric perspective, for the simple reason that one cannot see one’s egocentric perspective. One’s egocentric perspective is something only available to thought or imagination. Someone with his back onto the fire-place can imagine how things would look were he to face the fire-place.<sup>24</sup> One can move, change egocentric perspective at time  $t + 1$  and see the egocentric perspective one used to occupy at  $t$ . But one cannot see at  $t$  one’s current egocentric perspective at  $t$ .

The shift from the thought involving the binary relation to the thought involving the ternary relation illustrates a fundamental aspect of the cognitive dynamics inherent to conceptual representations. Notice that cognitive dynamics involves both the process of disengagement, whereby one ascends to a more detached thought, and the converse process of immersion, whereby one engages into a less detached thought.<sup>25</sup>

As noticed by Perry (1986b: 221), creatures with the conceptual power to form thoughts can use an  $n$ -place predicate in order to represent an  $n + 1$ -ary relation. If they do, then they have the power to move from a thought involving the  $n$ -place relation to a thought involving the  $n + 1$ -ary relation (i.e a relation taking  $n + 1$  arguments). Only conceptual thought can increase the arity of a predicate (i.e the number of its admissible predicates). Following Dokic (2002), we shall call this constraint the detachment constraint:

Creatures with the conceptual power to represent an  $n + 1$ -ary relation by means of an  $n$ -place relation have the power to move from a thought involving the  $n$ -place relation to a thought involving the  $n + 1$ -ary relation.

As the previous example illustrates, unlike conceptual thoughts, visual percepts do not satisfy the detachment constraint. The detachment constraint derives from the generality constraint or the language of thought hypothesis. Thoughts satisfy the detachment constraint because a thought is a structured combination of concepts. Thus, any conceptual parameter implicit in one thought can be made explicit in some further thought.

## 4.2 Unarticulated constituency

Clearly then, the process of cognitive dynamics is not limited to the transformation of visual percepts into thoughts. It involves shifts from more engaged to more detached or less engaged thoughts and vice versa, as a famous example of Perry’s (1986b) will make clear. Let us compare the thoughts expressed, respectively, by the three following sentences all uttered, e.g. in Lyon:

1. It is raining.
2. It is raining here.
3. It is raining in Lyon.

<sup>24</sup>In Chapter 3, Section 5.3, we will discuss neglect patients whose ability to imagine spatial perspectives is impaired.

<sup>25</sup>A point emphasized by Recanati (1997).

The meteorological thought expressed by an utterance of (3) contains a city name. Arguably, the meteorological thought involving a city name is more detached than either the thought involving the indexical ‘here’ or no expression for a place. Using a city name involves the ability to categorize spatial regions in terms of different contrastive cities. It may rain in Lyon, not in London. Furthermore, the use of the proper name ‘Lyon’ does not mandate a perceptual way of representing the location. Presumably, one may visually perceive parts of the city of Lyon, but one cannot visually perceive the city *per se*—except perhaps either as a dot on a map or from an airplane. But it is unlikely that one can see rain in Lyon either by seeing a dot on a map or by seeing Lyon from an airplane.

The meteorological thought expressed by an utterance of (2) in Lyon involves the spatial indexical concept expressed by ‘here’. The ability to identify a location using indexical or demonstrative concepts involves the ability to form perception-based thoughts about locations. Such thoughts are ‘egocentric’ in the sense that a location can be referred to as ‘here’ only from some agent’s subjective perspective, i.e. relative to the agent’s current location. The egocentric contrast between ‘here’ and ‘there’ is more engaged or less detached than the spatial contrast drawn by using a city name. One can form the thought expressed by an utterance of (3) while being, e.g. in New York; but one cannot form the thought that it is raining in Lyon by uttering sentence (2) unless one is located in Lyon. Hence, the meteorological thought expressed by an utterance of (2) is more narrowly tied or anchored to the perceptual context than the meteorological thought expressed by an utterance of (3).

Now, the question arises whether an utterance of (1) can express a thought at all, i.e. something that can be true or false. It rains (or snows) at some time, not others. If and when it does, it rains over some area, not others. No genuine disagreement can arise about whether a meteorological thought is true or false until the time and the place of the relevant meteorological event or state of affairs are fixed. Nor can an utterance of (1) be taken to express the thought that it is raining somewhere or other. Assuming that the temporal component of the meteorological thought is expressed by the present tense in (1)–(3), the location is explicitly referred to by a prepositional phrase in (2) and (3), not in (1).

Perry (1986b) argues convincingly that the thought expressed by an utterance of (1) in Lyon involves a tacit reference to some region of space that is supplied by the very fact that the sentence is uttered in Lyon. In Perry’s (1986b) terminology, (1) involves an ‘unarticulated constituent’ for the location: it is anchored in Lyon, not elsewhere. An agent looks out the window and comes to form the meteorological thought expressed by (1). This thought interacts with her action plans to generate her intention to, e.g. grab her umbrella. The agent did not bother to represent in her meteorological thought any contrast between the place relevant to her intended action and some different location—as she would had she instead formed the thought either expressed by (2) or by (3). In forming the thought expressed by (1), the agent lets the very area to which her immediate perception and action are anchored supply the missing location.

Perry (1986b) asks us to consider creatures from an imaginary country, which he calls ‘Z-land’. Z-landers, as he calls them, never travel and they are so fixated onto their land that their meteorological thoughts never contain a constituent for a place. Unlike normal human adults’ thoughts, the meteorological thoughts of Z-landers contain

a one-place predicate ‘rain’ with one argument for time, none for place. All the meteorological thoughts that Z-landers ever entertain are spatially anchored to their immediate surroundings, i.e. to the spatial context that is both perceptually accessible and relevant to their immediate actions. When Z-landers think what they express with their use of (1), the semantic connection between their thought and Z-land is supplied by their being in Z-land; but it is not explicitly reflected in their thought. As Perry (1986b) remarks, ‘there is a little of the Z-lander in the most-travelled of us’. We adult human beings, however, do have the ability to rise above the limitations of our visual perceptual capacities and reflect on the contrast between the region of space to which our actions and perceptions may be anchored, and other possible places affording different possibilities for perception and action.

Spatial demonstrative and indexical concepts (expressed by ‘here’ and ‘there’) lie at the very interface between the visual perception of spatial relations and spatial thoughts.<sup>26</sup> They afford elementary reflective conceptual capacities, which allow normal human adults to rise above the limitations inherent to Z-landers’ meteorological thoughts. As we argued above, the spatial egocentric contrast between ‘here’ and ‘there’ is more engaged than one made using a city name, but it is a conceptual contrast nonetheless. What shows that indexical and demonstrative expressions like ‘here’ and ‘there’ have conceptual contents is that they have inferential roles: if some object is *here*, then it is *not there*. Far more sophisticated is the human reflective conceptual ability to represent e.g. the fact that time is relative to a time zone. When a person living in Paris wakes up in the morning upon hearing her alarm clock, she does not have to represent the fact that it is 7:00 a.m. in Paris, not in New York; but she could. It was a considerable conceptual revolution to discover that simultaneity is a ternary relation. Rarely if ever, if they are not theoretical physicists, do human beings represent the fact that two events are simultaneous within one frame of reference, not others. But by learning special relativity, they can come to reflect on the fact that simultaneity is a ternary relation involving a pair of events and a frame of reference, not just a pair of events. In all such cases, the mapping of one thought to the next satisfies the detachment constraint: by a process of cognitive disengagement, one conceptual constituent implicit in one thought is made explicit in another thought.

## 5 Actions and the intentionality of intentions

Unlike plants, animals are agents, at least some of the time. Since they do not have roots, they must act in order to find food, protection, shelter and sexual partners.<sup>27</sup> Humans are no exception. Unlike plants, animals can act because they have a nervous system made up of neurons, some of which are sensory neurons, some are motor neurons and some are sensorimotor neurons. When an animal is an agent, then, unlike the motion of other physical objects, its action may involve, in Premack’s (1990) terms, a ‘self-propelled’ movement, i.e. a movement one of whose causes is internal to the agent.

<sup>26</sup>In Chapter 6, Section 7.2., we shall argue that what Campbell (1994) calls ‘causal indexicals’ lie at the interface between visuomotor representations and action concepts involved in the contents of prior intentions.

<sup>27</sup>Sponges, oysters, mussels, fetuses might be exceptions.



What is an action? The question is very complex. It has been, and still is, intensely discussed among philosophers. Although there is no settled answer to the question, we shall assume that an action is a special kind of behavior: intentional behavior. Much behavior is not intentional. Although plants do not navigate through space, let alone act, they nonetheless behave. Photosynthesis, e.g. is plant behavior. Deciduous trees shed their leaves in the Fall and flowers blossom in the Spring. Plant behavior is non-intentional. Much of what human and non-human animals do is non-intentional behavior too. Humans breathe, digest, salivate, blush, yawn, shiver, perspire, hiccup, snore, vomit and so on and so forth. If and when they engage in such behaviors, humans do not act. Reflexive behaviors are not actions. The light pupillary reflex produces the contraction of one's pupil. It is not an action. Nor is the visual control of one's posture.

A system *S*'s behavior is something the system does, as opposed to something that happens to *S*. If a tree loses its leaves because an animal has cut them or as a result of a storm, then losing its leaves is something that happened to the tree. Unlike shedding its leaves, it is not something the tree did. Being bitten by a mosquito is something that may happen to a mammal: it is a piece of mosquito behavior, not mammal behavior. If John raises his hand, then the movement of John's hand is part of his behavior. But if Ann raises John's hand, then the movement of John's hand is part of Ann's behavior, not John's. Of course, a system *S*'s motor output may have many different causes. But unless some internal state *c* of *S* did contribute to the process whereby *S*'s motor output *m* was produced, the motor output *m* was not part of *S*'s behavior.

On the face of it, however, *S*'s motor output *m* does not seem to be a necessary condition of *S*'s behavior. First of all, changes other than movements of an animal's limbs (or other bodily parts) may be involved in behavior. Humans, e.g. blush in certain circumstances. Blushing involves a change in the color of the face, not a movement of the limbs. Perhaps it is controversial whether blushing is something one does rather than something that happens to one. Concentrating and/or listening, however, are uncontroversially things one does and they do not involve specific movements of one's limbs. Second, behavior may consist in the lack of limb movement: a prey's behavior may consist in refraining from moving in order to escape its predator's attention. In this case, the prey's lack of movement is indeed controlled by one of its internal states. Producing a lack of movement is different from failing to produce any movement. The latter, not the former, is what a dead animal does—if a dead animal can do anything.

So behavior can be intentional or non-intentional. *S*'s behavior is an action if it is intentional behavior. What makes *S*'s behavior intentional is the peculiar nature of *c*, the internal cause of *S*'s movement (change or internally produced lack of movement). For *S*'s behavior to be intentional and hence to be an action, the internal cause of *S*'s motor output must be a mental representation. As Fodor (1986: 6) puts it:

[...] some of a paramecium's movements count as 'behaviors' in some sense of that notion that brings them within the explanatory purview of psychological theory. This is to grant that they count as responses but not, of course, that they count as actions. I take it only intentional systems can act.

## 5.1 The distinctive role of intentions in the etiology of actions

Prior to Davidson's seminal (1963) paper on 'Actions, reasons and causes', many philosophers were inclined to make a sharp distinction between reason explanations and

causal explanations. According to this distinction, it is one thing to provide reasons for an agent's intentional action, it is something else to supply the causes of an agent's physical movements. On this view, actions have reasons, not causes. Movements have causes, not reasons. Beliefs and desires can only be reasons for an action, not causes of physical movements. Davidson (1963) argued that an appropriate belief–desire pair can both be a reason for an agent's action and the cause of what the agent does. Subsequently, the question arose whether an agent is required to hold beliefs and desires in order for his behavior to qualify as an action.<sup>28</sup>

First of all, the question arises whether all actions must involve a relevant belief–desire pair. When it is spontaneously performed, or perhaps triggered by an affordance (in Gibson's sense),<sup>29</sup> an action may involve movements that are controlled neither by the agent's beliefs nor by his desires. Still we do not want to count all of them as reflexes.

Among acts that are not reflexes, there is the category that O'Shaughnessy (1980: 58–73) has labeled 'sub-intentional acts', i.e. a class of acts which are intentional under no description. Consider the movement of one's tongue in one's mouth as one reads, the movements of one's fingers as one is attending to a lecture or the tapping of one's feet to the rhythm as one is listening to a piece of music. Arguably, these acts are not instances of reflexive responses to an incoming signal. After all, one can stop to tap one's feet to the rhythm if requested to do so. But one cannot prevent one's pupil from contracting in the light or from dilating in the dark. Nor are these 'sub-intentional acts' intended in the sense that the movements involved in such acts can be causally traced back to an intention of the agent. Arguably, movements involved in 'sub-intentional acts' can be overt or covert: the movements can be executed or imagined. In this respect, the phenomenon discussed by O'Shaughnessy is interestingly related to what cognitive scientists call motor imagery.<sup>30</sup>

As emphasized by O'Shaughnessy (1980: 60), one can come to notice, and thus discover, the existence of such movements. What one so discovers then is the existence of bodily movements for which one is, as he puts it, 'responsible'. One discovers that one is executing or performing such movements. By contrast, the reflexive contraction of one's pupil or the visual control of one's posture are automatic responses to incoming signals for which one does not bear responsibility.

Also, as noticed by Searle (1983: 84), there seems to exist a class of intentional though *non-deliberate* actions: getting up and pacing about a room absorbed in one's thoughts seems to qualify as such a non-deliberate action. Again, the reflexive contraction of one's pupil or the automatic visual control of one's posture do not seem to qualify as non-deliberate intentional actions. Such non-deliberate actions differ from both reflexive behaviors and sub-intentional acts, for they do involve an intention on the part of the agent. Arguably, one cannot get up and pace around one's room without both intending and having decided to do so. One cannot, however, intend, let alone decide, to contract one's pupil. The contraction of one's pupil is the automatic result of the light pupillary

<sup>28</sup>See Pacherie (2001) for a detailed discussion.

<sup>29</sup>See Chapter 6, Section 2.

<sup>30</sup>See Chapter 6 for a discussion of motor imagery and Chapter 7 for a discussion of mirror neurons related to O'Shaughnessy's 'sub-intentional acts'.

reflex. Presumably, one does not intend, let alone decide, to move one's tongue in one's mouth while reading. Nor must one intend and decide to tap one's feet to the rhythm in order to do so while listening to a piece of music. One may, however, stop doing so.

On the taxonomy espoused here, one should distinguish reflexes from 'sub-intentional acts'. Although neither reflexes nor sub-intentional acts are intended, unlike the former, one can stop the latter if one intends to. Both reflexes and sub-intentional acts should be distinguished from non-deliberate intentional actions such as pacing around one's room, the existence of which shows that having beliefs and desires is not necessary for acting intentionally. Some intentional actions seem to involve movements caused by an intention alone.

Second, examples of so-called 'deviant' causal chains show that the contribution of a relevant belief–desire pair may not be sufficient for a behavioral process to count as an intentional action. Consider Davidson's (1980: 79) much discussed climber example:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold. Yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally.

The climber may feel the urge to loosen his grip on the rope. He may well know that if he did, he would be relieved. But if he happens to loosen his grip without ever intending to do so, then he did it by accident, not intentionally.

Searle (1983: 82) discusses another murderous example: Bill wants to kill his uncle because he believes that by doing so, he will inherit his fortune—something that would please him very much. He is driving his car towards his uncle's house thinking about how to kill him. On his way, he is so agitated by his thoughts that he accidentally hits and kills a pedestrian, who turns out to be his uncle. Although Bill killed his uncle, it was an accident—not something he did intentionally. Given his belief and desire, Bill did intend to kill his uncle, but *not* the way he actually did it. He did it by accident. Bill had what Searle (1983, 2001) calls a *prior intention* to kill his uncle. What he did not have is what Searle (1983, 2001) calls the *intention in action* to kill him by running him over with his car. Arguably, not all actions have prior intentions, but all have an intention in action. Getting up and pacing about a room may not be caused by a prior intention. But it must be caused by an intention in action.

Arguably, no single mental representation having the characteristic intentionality of intentions can play all the functions assigned to intentions in the etiology of actions. Thus, like many theorists who have embraced a dual view of intentions, Searle (1983) distinguishes between prior intentions and intentions in action. In his view, whereas the former represents the whole action, the latter represents the agent's physical movements necessary to achieve the action's goal. As Searle (1983: 93, 2001: 47–9) sees it, the prior intention results from a process of deliberation or reflection on one's own beliefs and desires. It in turn gives rise to the action, which consists of two components: the intention in action and the bodily movement. According to Searle's (1983, 2001) particular view of the matter, the prior intention would seem to stand as the cause of both the whole action and the intention in action. The prior intention would stand as a cause of the whole action by virtue of being the cause of the intention in action, which in turn

causes the bodily movement. If so, then it seems to be a consequence of Searle's view that, unlike the intention in action, the prior intention does not belong to the action properly so called: since it causes it, it must be distinct from it. Searle's dual view gives rise to the further question whether the prior intention fades away once the intention in action is formed, as it seems it must once its causal role is achieved. If and when a short-circuit causes a fire, once the fire is on, the short-circuit is over. Similarly, if a prior intention causes an intention in action, the prior intention should give way to the intention in action.

## 5.2 The intentionality of intentions

Perceptions, beliefs, desires and intentions are psychological states with intentionality: they are about or represent objects and states of affairs under a particular psychological mode or format. Perceptions, beliefs, desires and intentions each have a distinctive intentionality. Anscombe (1957: 56) asks us to consider a 'shopping list'. The list might either be used as a set of instructions (or a blueprint) for action by a customer in a store or as an inventory by a detective whose purpose is to draw a record of what the customer is buying. In the former case, the list is not to be revised in the light of what lies in the customer's grocery bag; but in the latter case, it is. If a mismatch should occur between the content of the grocery bag and the list used by the customer, then the blame should be put on the customer, not on the list. In the case of a mismatch between the content of the bag and the list drawn by the detective, the detective should correct his list. In Searle's (1983, 2001) terminology, beliefs and desires have opposite 'directions of fit'. Beliefs have a mind-to-world direction of fit: they can be true or false. A belief is true if, and only if, the world is as the belief represents it to be. It is the function of beliefs to match facts or actual states of affairs. In forming a belief, it is up for the mind to meet the demands of the world. Unlike beliefs, desires have a world-to-mind direction of fit. Desires are not either true or false: they are fulfilled or frustrated. The job of a desire is not to represent the world as it is but rather as the agent would like it to be. Desires are representations of goals, i.e. possible non-actual states of affairs. In entertaining a desire, it is so to speak up for the world to meet the demands of the mind. The agent's action is supposed to bridge the gap between the mind's goal and the world.

As Searle (1983, 2001) has noticed, perceptual experiences and intentions have opposite directions of fit. Perceptual experiences have the same mind-to-world direction of fit as beliefs. Intentions have the same world-to-mind direction of fit as desires. In addition, perceptual experiences and intentions have opposite directions of causation: whereas a perceptual experience represents the state of affairs that causes it, an intention causes the state of affairs that it represents.

Although intentions and desires share the same world-to-mind direction of fit, intentions are different from desires in a number of important respects, which all flow from the peculiar commitment to action of intentions. Broadly speaking, desires are relevant to the process of deliberation that precedes one's engagement into a course of action. Once an intention is formed, however, the process of deliberation comes to an end. To intend is to have made up one's mind about whether to act. Once an intention is formed, one has taken the decision whether to act. Of course, every intention does not give rise to an action. There are more intended actions than actions performed, for the simple

reason that human beings change their minds. We shall discuss the contrast between intentions and desires under five headings.<sup>31</sup>

First, as we said above, intentions, unlike desires, achieve their world-to-mind direction of fit by virtue of their mind-to-world direction of causation. This is what Searle (1983, 2001) calls the ‘causal self-referentiality’ of intentions. If, e.g. one intends to raise one’s arm, then one’s arm must rise as a result of one’s intention to raise it and not otherwise. If Ann raises John’s arm, then the rising of John’s arm cannot result from John’s intention. John may have wished that Ann would raise his hand, but John cannot intend that Ann raises his arm. In our view, what Searle (*ibid.*) calls the causal self-referentiality of intentions does not require that only creatures with the conceptual resources necessary for representing the causal relation between their intentions and their subsequent movements can be agents and be ascribed intentions.

Although the issue is complex, we would like to suggest that what Searle characterizes in terms of causal self-referentiality is a feature of intending, i.e. the psychological mode of intentions. Assuming that the human cognitive architecture includes an ‘intention box’, as it includes a ‘belief box’ and a ‘desire box’, only some mental representations have the appropriate format for entering each of these ‘boxes’.<sup>32</sup>

Thus, when an agent forms an intention, the content of her mental representation is entertained under a mode distinct from both beliefs and desires, whether the agent knows it or conceptualizes it or not. The agent need have no higher order belief about what it takes to form an intention, as opposed to a perception, a belief or a desire. Perhaps, as discussed by Searle (1983, 2001) and by Pacherie (2000b, 2001), what William James called ‘the experience of acting’ is the non-conceptual phenomenological counterpart of what Searle conceptualizes as the causal self-referentiality of intentions. A patient whose arm has been anesthetized is requested to raise his arm with his eyes closed. Unbeknownst to him, his arm is prevented from moving by being tied to his chair. Although he cannot raise his arm, he has the experience of acting, so much so that, upon opening his eyes, he is surprised to see that his arm did not rise. Conversely, epileptic patients have been stimulated by Penfield (1975) by direct application of a microelectrode onto their motor cortex. The stimulation of the motor cortex caused the movement of the patient’s arm. The patient, however, was surprised to see his arm rise and reported no experience of acting. Perhaps James’ experience of acting is related to both the sense of agency and to what is called ‘motor imagery’ (see Jeannerod 1994, 1997).<sup>33</sup>

Second, as already alluded to previously, desires can be about anything or anybody: one can wish or hope that somebody else would do something or other. Unlike hopes and desires, however, intentions are always about the self. One intends to raise one’s arm, not somebody else’s. One intends to reach a glass of water, to get married or to get elected. The intended reaching, marriage or election must be the agent’s own. One can wish somebody else would reach a glass of water, would get married or would get elected. One can intend to contribute to the reaching of a glass of water by somebody else. One can intend to contribute to somebody else’s marriage or to somebody

<sup>31</sup>Millikan (1996) argues that human intentions are hybrid mental representations with both a directive world-to-mind direction of fit and a descriptive mind-to-world direction of fit. In Chapter 6, Section 8.1, we discuss her view.

<sup>32</sup>The functional terminology of ‘boxes’ was introduced by Schiffer (1981) and Fodor (1987).

<sup>33</sup>See Chapter 6, Section 8.1 for further discussion of motor imagery.

else's election. But one cannot intend that somebody else reaches for a glass of water. Nor can one intend that somebody else gets married or gets elected. Similarly, a human being can wish she were, e.g. a bird; she can pretend, but she cannot intend, to be a bird.

Third, unlike desires, intentions are tied to the present and to the future. One cannot form an intention about the past. One can wish things had been different in the past, but one cannot intend to have done something in the past. Intentions can be about temporally distant states of affairs lying far ahead in the future or about immediately executable goals. The more the state of affairs intended is temporally remote from the time when the intention is being entertained, the more the content of the intention is conceptual (or conceptualized). Hence, intentions directed towards states of affairs remote in the distant future may and must have conceptual content. The more the goal is accessible temporally and spatially, the less it needs to be conceptualized. The conceptual content of intentions about remote goals involves action concepts (with a world-to-mind direction of fit).

Fourth, like the contents of desires, the contents of intentions are about possible non-actual states of affairs. But unlike the content of a desire, the conceptual content of an intention directed towards the distant future cannot be about a state of affairs that the agent knows to be impossible. An agent may wrongly take a state of affairs to be possible. Contrary to her expectation, the intended state of affairs may turn out to be impossible. But an agent cannot intend to achieve a state of affairs that she knows to be impossible at the time when she forms the intention.

Finally, one can consistently entertain contradictory desires, but one cannot consistently form contradictory intentions. One can consistently have the desire to be at the same time in Paris and in New York city: one can consistently wish one were simultaneously in Paris and in New York city. A man can consistently have the desire to marry two distinct women at the same time; he may consistently wish to do so. But one cannot consistently form intentions that cannot be carried out simultaneously. One can consistently wish to be, but one cannot consistently intend to be, at the same time in Paris and in New York city. Although he can consistently wish he could get married to two distinct women simultaneously, nonetheless, in many human cultures, a man cannot consistently intend to marry two different women at the same time.

In the rest of this book, we shall review evidence in favor of the hypothesis that the human visual system processes visual information about objects in two fundamentally different ways: one and the same stimulus can undergo perceptual processing or motor processing. For example, one can perceive a cup and one can grasp it. We do not want to suggest that the human visual system is restricted to perceiving objects of prehension: humans can perceive a great many things other than objects that can be direct targets of their hand actions. Nor do we want to suggest that humans can only plan and intend object-oriented actions: humans can intend actions far more complex than grasping an object. In this chapter, we have laid the groundwork for the rest of the argument by emphasizing the contrast between the mind-to-world direction of fit of both beliefs and perceptual experiences, and the world-to-mind direction of fit and the mind-to-world direction of causation of intentions. As we shall argue in Chapters 5 and 6, the output of the perceptual processing of visual inputs serves as an input for further conceptual processing that produces thoughts about objects, which can be stored in the 'belief box'. In Chapter 6, we shall argue that visuomotor representations, which result from

motor processing, present visual information about objects to motor intentions. Notice that in our view, visuomotor representations interact with motor intentions. They are not involved in what O'Shaughnessy (1980) calls 'sub-intentional acts', i.e. bodily movements whose causal origins cannot be traced back to an intention of the agent. Although we do believe that 'sub-intentional acts' in O'Shaughnessy's sense do indeed exist, we do not believe that visually guided hand actions directed towards objects are 'sub-intentional acts' in O'Shaughnessy's sense. Visuomotor representations are hybrid representations with a dual direction of fit, or so we shall contend: they represent features of actual states of affairs appropriate for action. We shall further argue that the contents of visuomotor representations can be conceptualized with the help of action concepts and be stored in the 'intention box'. Thus, in the rest of this book, we shall claim that the duality between the perceptual processing and the motor processing of visual objects is the reflection within the human visual system of the duality between the intentionality of intentions and the intentionality of perceptions and beliefs. In a few words: one can intend to act, but one cannot intend to think (or believe), let alone to perceive.