MONASH University

Perceptual phenomenology and predictive processing

Ryoji Sato
M.A.

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2016
Philosophy Department

Copyright notice

Abstract

This thesis is about five aspects of our perceptual phenomenology that all coincidentally start with P: presence, poverty, present, particularity, and persons.

All of these are much discussed in contemporary philosophical debates but they are also notoriously difficult to explain. The thesis is an attempt to give new conceptual and phenomenological analyses of the five aspects and identify common core features among all of them. It turns out that they naturally invite a contemporary neurocomputational explanatory framework: predictive processing. In this way, the thesis provides a novel, unified explanatory approach to distinctive aspects of human perceptual world.

Despite its diversity of topics, a common theme emerges after conceptual clarification. The topics are, at their core, all related to the topic of how representastions of deep aspects of the world and those of more shallow, palpable aspects of the world interact and how they form coherent percepts. This calls for a hierarchical structure equipped with precision expectations in predictive processing. By appealing to predictive processing, it is shown that higher level models "create" contents at lower levels, where relevant sensory stimulation were not available at the time of experience.

The approach of the thesis is interdisciplinary—I use both philosphy and cognitive neuroscience to elucidate the phenomena, All the chapters for the five Ps begin with philosophical analysis of the subject matter, which then sets the scene for, and is facilitated by, the predictive processing framework. In turn, science can benefit from philosophical analysis and I formulated empirical predictions in each chapter.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: [redacted]

Print Name: …Ryoji…Sato………

Date: ……16/12/2016………….

Acknowledgements

I first would like to thank my supervisors: Prof. Jakob Hohwy and Dr. Jennifer Windt for sparing a lot of time at the end of year. I would like to thank Prof. Hohwy for his extreme enthusiasm and creativity. He has always given me constructive, and eye-opening feedbacks underwritten by immense knowledge both in philosophy and cognitive science. I would also like to thank Dr. Windt for her kindness and carefulness. She substantially raised the quality of my thesis through her encouraging comments and careful feedbacks. The period after Jennifer joined my supervisory team was the most product period during my candidature. My gratitude also goes to Dr. Monima Chadha, for her consistent kindness and insightful comments. I would like to thank Lucy Mayne, who has proofread and edited my thesis, for her incredible knowledge about proofreading and editing and her kindness to devote her time for my thesis at the busy time of year. Last but not least, I would like to thank my family, for their continuous support throughout my prolonged studentship.

Table of Contents

1. Introducing the Five Ps

1.1. *Our complex perceptual phenomenology and the brain: an illustration from the Game of Life*

This thesis is about five aspects of our perceptual phenomenology: presence, poverty, present, particularity, and persons. These perceptual aspects are much discussed but remain conceptually evasive and difficult to explain. The thesis provides new analyses of the five aspects, and identifies a common thread through all of them. This motivates a new and unified approach to them, conceived through a contemporary neurocomputational explanatory framework. The thesis thus provides a new, unified and explanatory approach to distinctive aspects of how we perceive the world.

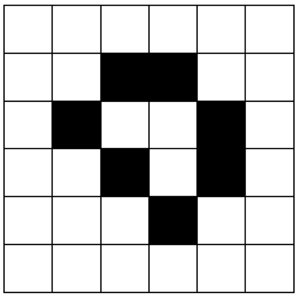
To set the scene, I first consider a simple type of artificial system with only rudimentary perception. In the early history of research on Artificial Life (and in the infancy of computer science), Conway's *Game of Life* was one of the first simulators of life, and was widely popular because of its simplicity and the intriguing patterns it produced.[1] The game involves a grid with a number of cells that only have two possible states—one corresponds to life and the other corresponds to death. An individual cell changes its state following simple rules that only concerns the state of eight adjacent cells, called 'neighbors'. There are only four rules governing life and death of each cell:

> Reproduction: If a cell is dead and has exactly three live neighbors, the dead cell springs into life.
> Underpopulation: If a cell is alive and surrounded by fewer than two live neighbors, it dies.
> Overpopulation: If a cell is alive and surrounded by more than three live neighbors it dies.
> Survival: If a cell is alive and surrounded by two or three live neighbors, it survives to the next generation.

[1] The Wikipedia entry for Conway's Game of Life (https://en.wikipedia.org/wiki/Conway%27s_Game_of_Life) contains some illustrative animations. There are also some websites where you can play the game of life for yourself. For example, go to https://bitstorm.org/gameoflife/.

A cell is born when surrounded by the right number of kin and it dies when it is surrounded by too many or too few of them. Although these rules are caricatures of the behavior of biological systems, they nevertheless seem able to produce life-like patterns. The patterns created look like the rise and fall of simple creatures—colonies of bacteria, perhaps. A colony is moving in one direction, bumps into another colony, they interact, divide into two colonies again and so on. There are some intriguing patterns, which eventually acquired particular names. Some are staying alive forever ("still life"), some repeat fixed patterns ("oscillator"), some travel across the grid ("spaceship"). (See Figure 1 below for some examples).

Sill Life



Oscillator (Cycle = 2 unit time)

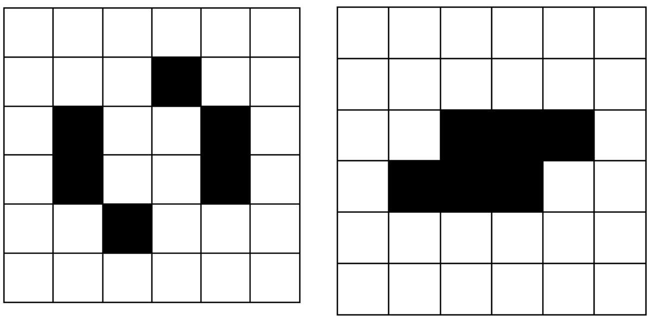


Figure 1 Some patterns in the Game of Life

Now, consider the perspective of one living cell in the Game of Life—it "perceives" the states of the eight cells surrounding it and "decides" what to do next. What it sees is only the states of the eight adjacent cells at a specific time—what a simple perceptual world compared to ours! It is simple in many senses. For example, for this simple artificial life, content is relatively informationally poor. Its perception only gives information about the on-or-off state of eight cells—only eight bits of information. Its perception also lacks the *structure* that human conscious perception has. For example, the life doesn't have perspectival appearance, its sensorium is the same from all perspectives. All its perception concerns is presence or absence of life. In contrast, we have only limited perspectives of an object at any given time. When you look at a coffee mug, there are always currently unseen sides of the mug despite the fact you perceive it as a mug rather than as an aspect of a mug. Moreover, its perception concerns states of the adjacent cells only at a specific time, thus persistence of cells throughout time is not considered at all. Hence, there could be no perception of motion by a cell because motion takes place over time. From the outside, we do perceive motion in the Game of Life, but the life in there does not perceive any time-extended events; what happens at any given time is not remembered, and no prediction of the future is made. Both of these structural and informational aspects of human conscious perception help set the main agenda of this thesis: the thesis is about the deeper, more subtle aspects of perceptual phenomenology.

We can modify the Game of Life so as to make the life in the game look more realistic, while retaining the simplicity. We might add predators, prey, and simple rules that govern the behavior of each kind of entity. For example, if a live cell is surrounded by three or more predators, it will be dead in the next generation ("predation",) or if a live cell is surrounded by two predators or fewer, it survives by moving to a cell that is not occupied by a predator ("flee".) This imaginary creature looks more like a real living creature—but the structure and content of its perception remains much more simple than ours: it just detects the presence or absence of food predators, and members of its kind at that particular instant. Let us call this creature Simpleton and its experiences Simpleton's experiences.

We can also talk about real creatures that have much simpler perception in the sense discussed above. Paramecia, actual biological systems, seem close to Simpletons. When a paramecium bumps into an obstacle, it senses the object with its cilia and turns in the other direction. It also moves toward slightly acidic environments, because that is where it tends to find its food, bacteria. The paramecium's perceived world seems not so different to that of the Simpleton. It might represent "There is an obstacle!", "In this direction, food!", or "Here, food!" Its perceptual states are informationally poorer and less structured than ours. It might not represent perspectival appearance of objects and it might not represent time-extended events.[2]

We can climb further up the ladder of evolution, going to more complex creatures. Think about a frog. It is well known that it captures any small flying black object detected in its visual field, whether it is a real fly or a pellet. Its perception is not as simple as that of Simpleton or paramecia. Since it detects flying black objects, it might represent the size or the color of the object. It probably also represents the *motion* of the object (at least to distinguish it from sensory change due to the frog's own movement). In this sense, the perception of a frog is structurally richer: it can represent time-extended events. Thus, it would represent a black object flying in a specific direction, and on top of that, it might also represent the object as food although it cannot distinguish between flies (real food) and other black flying objects (as Millikan (1989) has argued). However, the frog would not represent the objects with further details. It does not, let us assume, represent any distinguishing characteristics between flies and other black flying objects, such as their particular shape and the noise they make, as it does not seem to rely on such differences to avoid errors in its food-seeking behavior. The frog's perception also would not concern the particular identity of the food item. It wouldn't represent whether the fly is the same one as the one flying yesterday, for example. The frog's perceptual experience would be closer to our perceptual experience than to that of the Simpleton and paramecium, but it still lacks salient aspects of the perceptual phenomenology of human beings.

Compared to the perception of those creatures, we human beings enjoy quite rich and structured perceptual experience in daily life. Sometimes in philosophy, in an effort to make things clear and simple, there is a tendency to consider examples as simple as that of the Simpleton. But all the aspects of perceptual experience I discuss in my thesis are those structural or rich aspects of human perceptual experience. We can use the experience of the Simpleton as a contrast case to understand the uniqueness and

[2] It is ultimately an empirical matter whether paramecia represent the perspectival appearance of an object or time extended events. However, it seems that they have to be able to respond differently to different aspects of objects to represent perspectival appearance and there needs to be evidence that paramecia remember the recent past of an object and change their behavior on the basis of memory and prediction from that to ascribe, say, the representation of motion of an object.

significance of the aspects of human perceptual experience I am going to discuss.

Consider this intuitive example first. Suppose you are, somewhat awestruck, looking at the famous big Buddha statue in Nara, Japan (see Figure 2 below). You see the statue from one perspective at a time; you cannot see the front and the back of the statue at the same time. Yet you feel the *presence* of the whole Buddha statue in perceptual experience. You also somehow sense the presence of the back of statue—even though you cannot see. This structural complexity is different from the experiences of Simpleton.

Figure 2 Buddha statue in Tōdai-ji Temple, Nara, Japan[3]

Also, facing the Buddha statue, you experience the intricate decorations

[3] Photo by (c)Tomo.Yun (http://www.yunphoto.net)

surrounding the statue and spanning all over your visual field—the big Buddha's halo, and the little golden Buddhas encircling him. Intuitively, the content of human visual experience is much *richer* than the perceptual experience of the Simpleton, which only carries information about presence or absence of two kinds of creature in the surrounding eight cells.

There are also aspects of perceptual experience that are essentially related to time. Just as a frog experiences motion, we also experience time-extended events. For example, we visually experience a bird's flying across our field of vision, or the tourists milling around the Buddha's statue. When we have a visual experience of a flying bird, this phenomenologically goes beyond a mere collection of discrete snapshots of the bird's position at different times; our experience of a flying bird seems to involve the smooth transition of the bird's position.

1.2. *Aspects of Phenomenology: the five Ps*

This thesis is an attempt to account for the structural aspects of our perceptual experience—the subtle, deep, rich aspects that go beyond the simple perceptual experiences of imaginary Simpletons, and those oftentimes studied in some traditions in the philosophy of perception. I will seek to formulate this account using an increasingly popular neurocomputational theory: the *predictive processing* framework. The thesis will focus on five aspects of perception, hinted at above in the discussion of what Simpleton's lack. Coincidentally, these aspects all start with 'P'[4].

1. Presence
2. Poverty
3. Present
4. Particularity
5. Persons

I shall refer to these aspects as "the five Ps". All of these aspects are hotly

[4] 'P' is somehow a popular letter, which shows up in many different disciplines. has an interesting paper on this titled: "Alliteration in medicine: a puzzling profusion of p's." (Hayden, 1999)

discussed both in philosophy and cognitive science. They are also closely related each other. Their deeper connections will be gradually revealed as we examine them through the lens of predictive processing. First, let me briefly introduce each of them. (The summary of five Ps is provided in the end of this section.)

The first P to be discussed is *Presence* (Chapter 3). Examples of perceptual presence range across a wide area, but a typical example is the perception of the Buddha statue, I described above. When you see an object, you not only experience the front of the object, but also experience it as the whole object. It sounds almost contradictory, but we somehow feel the presence of currently unseen sides of an object. The complex aspects of perception are the targets of the chapter on Presence.

The second P is *Poverty* (Chapter 4). Intuitively, it seems that our visual experiences contain rich details across the visual field. However, based on empirical findings, philosophers and scientists including Daniel Dennett (1991), and Alva Noë and Kevin O'Regan (2000) claim that this is an illusion: even though we intuitively think that we experience rich details throughout our visual field, we do not actually experience vivid details in our peripheral vision, or in the areas where we do not attend. In other words, our visual experience is more impoverished than we think; hence 'poverty'.

This claim rests on the assumption that we do in fact intuitively believe that we experience rich details throughout our visual field. However, Jonathan Cohen (2002) argues against this assumption. Thus, this chapter has a two-tier structure—the first part discusses introspection of perceptual experience and the other part discusses the nature of perceptual experience itself and examines how rich our perceptual experience is.

The third P is *Present* (Chapter 5). This chapter focuses on the temporal aspects of perceptual experience, also introduced above in the example of the perceptual experience of a bird's trajectory across the sky. In an important sense of the word, perceptual experience is as of the present. If you perceive an apple in front of you, the apple exists in front of you *now*. You can recall the apple you ate last night or anticipate the apple you intend to buy on your way home later, but this is not perception. Moreover, our perceptual experience has another important temporal feature: duration. This derives from the observation that we can perceptually experience time-extended events. We can see a bird's flying or listen to a dog's barking. Flying or barking essentially involves time-passing. For example, in the case of barking, if a dog barks “bow-wow,” there is a temporal order—“bow” is followed by “wow.” This is puzzling, because, intuitively, we can experience “bow-wow” now. However, it seems impossible to experience “bow-wow” *now* if there is a temporal order; when you are experiencing “bow” then you are not experiencing “wow”, and when you are experiencing “wow”, then you are not experiencing “bow”. In Chapter 4, I will give an account of this puzzling notion of the experienced present, often labeled the ‘specious present’. These temporal features of our perceptual experience make a stark contrast with perceptual experience of Simpleton. The Simpleton’s experience does have presentness, in some sense, as it is about the present states of adjacent cells (though it is not clear there would be any phenomenological difference to any remembered or anticipated states). However, its perception does not concern any time-extended events. It is not about the time-course of living cells, for example.

The Fourth P is *Particularity* (Chapter 6). Perception does not seem to be about random objects. It is rather about *particular* objects in the *vicinity* of the experiencer. If you experience an apple, the experience is about *that* apple rather than a random apple qualitatively identical to that apple. However, the claim that we experience particularity of objects per se is in tension with a simple thought experiment—if the apple you are experiencing is swapped, without your knowledge, with another sufficiently similar apple, you wouldn't notice the difference. This thought experiment suggests that perceptual experience is not sensitive to the particularity of objects after all. How, then, is our phenomenology of particularity explained? This chapter addresses this aspect of perceptual experience and the cognitive mechanisms that buttresses particularity. Is there any particularity of this sort in the perceptual experience of the Simpleton? The answer seems both yes and no. It has particularity in the sense that the perception is about individuated items nearby and these items are not conflated with one another (“That is a predator left of me”, for example). But the Simpleton’s perception does not track the object over time, which our perception does. We can perceptually track a fly over time and identity it despite its change in location if we see the fly continuously. These different aspects of particularity will also be discussed in the chapter.

The last P is *Persons* (Chapter 7). We encounter many people in daily life—family, friends, colleagues, shop staff, random pedestrians, etc. When it comes to familiar people, it seems that we can directly perceive their identity. When you see your close friend at a bar, your recognition of him or her is instant and has a perceptual feel (e.g., “I saw Joe”). Most of the time, you don’t need to consciously infer who the person is based on their physical properties, you just know. This aspect is more dramatically highlighted by the existence of disorders of person identification: misidentification syndromes. For example, patients with Capgras delusion, a version of misidentification, typically insist that a close friend or family member has been replaced by an imposter, despite that person not having changed in appearance. Capgras patients admit the similarity of the alleged imposter, yet they insist that the person before them is a different person from the familiar person. As we have seen, all five Ps involve aspects of perception unique to humans but they look diverse. What are their commonalities?

Table 1 The Five Ps

Aspects of Conscious Perception	**Brief description**
Presence	We do not experience only the seen sides of an object, we also experience the object as a whole. In particular, we experience unseen sides of an object.
Poverty	Intuitively, our visual experience is rich in detail throughout the visual field. But current empirical evidence suggests we have little information about our peripheral vision and unattended objects.
Present	Out perceptual experience is as of the present, but this present seems to have some duration (“specious present”).
Particularity	Phenomenologically, our perceptual experience seems to be of a particular object rather than of a random object that has certain properties.
Persons	Phenomenologically, we tend to experience particular familiar people directly without consciously identifying them from their characteristics.

1.3. *Structure and commonalities in the five Ps*

As I mentioned above, the five Ps are not just aggregates of different aspects of perceptual phenomenology, and they have a common theme beyond the fact that all are topics of intense discussion in philosophy of mind and cognitive science. They are often not characterized clearly and discussed more on a case-basis. Because of that, it is not easy to see the underlying commonality among the five Ps at the outset. However, after careful conceptual clarification, a common theme emerges.

> *The five Ps are, at their core, all related to the topic of how representations of deep aspects of the world and those of more shallow, palpable aspects of the world interact and how they form a coherent percept.*[5]

What are shallow and deep aspects of the world? An example might help here. In the chapter on persons we will discuss whether personal identity is perceivable. It is easy to see the shape of a person or the color of his or her clothes. Those properties are phenomenologically salient and the causal story from those properties to sensory organs seems straightforward.[6] In the visual case, light is reflected by the surface of an object and it hits the retina, and the reflected light contains information about shape and color. But what are the sensory differences a person generates in virtue of being a particular person? While each person has a specific face, voice, temper, etc, and these features causes idiosyncratic sensory differences, it is nevertheless difficult to see what combination of those features the brain relies on to establish personal identity.

As this example reveals, by deep aspects of the world, I mean aspects of the world that do not leave much immediate sensory evidence on us. And by shallow, palpable aspects of the world, I mean aspects of the world that are easily detectable in sensation. I will explain this distinction in more detail in Chapter 2. The topic of interaction between the deep and shallow aspects of the world is not as easily seen for all the Ps. However, as the discussion proceeds in each chapter, we will find a similar interplay between deeper levels and shallower levels.

Crucially, predictive processing is a theoretical framework particularly well

---

[5] In contrast with perceptual experience of the Simpleton, the five Ps are more "objective" aspects of the world in the sense it is not necessary tied to a certain kind of action. The content of the Simpleton's experience is more action-oriented.

[6] However, we will see this is actually not the case when I introduce of predictive processing in the next chapter (see Section 2.2.1, "*Prediction Error Minimization*").

equipped to do this excavation of the deep causal hierarchy. As we will see shortly, this framework postulates a hierarchical system in the brain that filters out different aspects of the world in a manner that fits with the shallow-deep distinction and which therefore presents itself as a candidate for explaining the five Ps in a unified manner.

The theme of how representations of the deeper aspects of the world are related to the shallower, and how they form a coherent percept, raises many related questions.

Firstly, this theme is related to the famous *inverse problem*, which has to do with the inference of the nature of a cause from its effect on the sensory organs; this is an inversion of the causal order, hence the name. The sensory stimulation registered to sensory organs at a time is insufficient for the brain to represent its causes. For example, sensory stimulation is similar both when you see a shadowed white surface and when you see a gray surface under good lighting conditions. The inverse problem affects all five aspects. For example, what the unseen back of an object looks like is underdetermined by the seen representation of the front of the object. In this case, there is no sensory stimulation coming from the back of an apple but we nevertheless somehow feel the presence of the back (this topic will be discussed in Chapter 3). This problem is ubiquitous in perception, including more conventional contents of perception such as the example of gray and white surfaces given above. The inverse problem calls for more top-down, inferential approaches to perception (e.g., knowing something about the context afforded by shadows and light). Predictive processing is a theory that relies heavily on expectations, and it thus provides a neat solution to the inverse problem and thereby for some of our Ps, as we will see in the following chapters.

Secondly, most of the five Ps are also subtle and elusive aspects of perceptual phenomenology; apt description of the phenomena itself is not an easy task. This gives rise to the *description problem.* For instance, I mentioned that we intuitively think that our perceptual experience is rich, but what does it mean for perceptual experience to be rich? Or when it is said that our perceptual phenomenology reflects ‘particular’ objects, what does this actually and precisely mean? Is it that the real particular objects in the world comprise our perceptual experience or is it just that some sort of directness is included in phenomenology? The description is so difficult that philosophers never seem to converge on any particular conception of any of the five Ps. Here, the predictive processing framework can contribute to the debate. The predictive processing framework stands alone as a neurocomputational theory but it also serves to constraint the description of perceptual phenomenology in the relevant chapters.

Thirdly and relatedly, because of its phenomenological subtlety and depth in the world, the perceptuality of the five Ps is often questioned, giving rise to the *perceptuality problem*. For example, Nanay (2010) argues that the representation of the back of an object is not perception but imagery (his argument will be discussed at length in Chapter 2). Similarly, I will discuss the representation of personal identity in Chapter 6, but someone might argue we do not *perceive* personal identity, rather, it is a *judgment* reached after the perception of the face or some other parts of the person. The argument is difficult to settle without a principled way to think about what perception is at heart. Predictive processing can be of use here, since it suggests one way to think about the nature of perception. In the final section of this chapter, I will give a working definition of perception conceived from the perspective of predictive processing.

The last problem is the *reason problem*. The five Ps are subtle, but concern the complex structure of conscious perception. But why does our conscious perception have this complex structure at all, rather than giving the simple experience of the imaginary Simpleton creature? This problem can be answered in many different ways; one obvious way is to appeal to evolution. For example, humans may have such complex perception because it was advantageous for survival. However, my approach here is different. I will instead show how the hierarchical Bayesian inference mechanism, which the predictive processing framework approximates, enables the five Ps. I will suggest why we have the hierarchical Bayesian inference mechanism in the first instance, and thus explain the occurrence of the five Ps.

The problems introduced above will be addressed in the subsequent chapters for each of the five aspects of perception, according to the following common structure:

> **Problem 1: Inverse problem**
> How can we experience aspects of conscious perception related to five Ps given that our sensory stimulation is so limited?
>
> **Problem 2: Description problem**
> Given that aspects of conscious perception related to five Ps are elusive, how can we describe phenomenology aptly?

Problem 3: Perceptuality problem

Is each aspect of conscious experience related to five Ps perceptual or cognitive?

Problem 4: Reason problem

Why do each of the five Ps appear in our conscious perception?

1.4. *The five Ps and predictive processing.*

As I have mentioned above, I will use the predictive processing framework to address each of these problems, and give an account for how this applies to each of the five Ps. Predictive processing is essentially a Bayesian approach to cognition. The predictive processing framework I rely on in this thesis is mainly developed by Karl Friston and his collaborators (Friston, 2010, 2011, 2012; Friston, Adams, Perrinet, & Breakspear, 2012; Friston, Thornton, & Clark, 2012; Mathys, Daunizeau, Friston, & Stephan, 2011). What separates this theory from other neurocomputational theories is its fundamentality and domain-generality. It is a theory of whole brain functioning that starts from the "free energy principle," which states any organisms that are in equilibrium with their environment have to minimize free energy. Predictive processing itself is a special case of the free energy principle that holds under certain conditions. I will explain free energy principle in detail next section. Because of this fundamentality and domain-generality, the predictive processing framework can provide a principled way to understand many aspects of the human mind. This characteristic of the predictive processing framework enables us to answer the description question, the perceptuality question, and the reason question.

We can now take a closer look at how accounts of the five Ps can benefit from using this theoretical framework. Let's start by acknowledging that all five Ps are *representational aspects* of conscious perception: all five Ps are about how different aspects of the world are represented by the brain. Presence is about how different aspects of an object are represented, Poverty is about how much of the world is typically represented, Present is whether and how the time-extended events in the world are represented, Particularity is about whether and how a particular object is represented, and Persons is about whether and how particular persons is represented. One important aspect of the predictive processing framework is that it is the theory of how the brain comes to represent the world on the face of ambiguous sensory stimulation (the inverse problem). Thus, assuming that predictive processing can indeed capture those mentioned representational aspects of the five Ps, we may plausibly appeal to this neurocomputational theory of representation.[7]

That said, predictive processing is a neurocomputational theory of the brain so it concerns representational properties *of the brain*. Thus, we have to make some metaphysical assumptions about the relationship between representational properties of the brain and representational properties of conscious perception to get the predictive processing account off the ground. Some philosophers are reluctant to investigate conscious perception using empirical theories. For example, Dainton (2006, p. *xv*) says, "As long as the matter-consciousness relationship remains problematic, the only way these questions can be answered is by inspecting consciousness itself, from the inside". I disagree. Even though the identity between the mind and the physical brain is far from established, some systematic relationships between matter and consciousness are relatively uncontroversial.

I propose that description of phenomenology and description of the cognitive architecture of the brain are mutually constrained. That is, the description of phenomenology has to be compatible with our description of brain function at many levels, and vice versa.

A good example of such mutual constraint comes from studies of binocular rivalry by Logothesis and colleagues (Logothetis, Leopold, & Sheinberg, 1996; Logothetis & Schall, 1989; Sheinberg & Logothetis, 1997). Binocular rivalry is a visual phenomenon that takes place when a subject is presented with two different images, one to each eye. The subject experiences alternating images rather than experiencing a fused image of the two (although fusion also happens occasionally). In Logothesis and colleagues' studies, a monkey is used as the test subject. The monkey is trained to press different bars when presented with different images, for example, faces and sunbursts. After the

[7] Since all the accounts of the five Ps are representational stories, success of the accounts does not depend on the hard problem of consciousness (Chalmers, 1996). The hard problem is the problem of why we have consciousness *at all*, whereas the accounts given in this thesis largely concern why we have conscious perception *of particular types of contents* given that we have consciousness.

training, the monkey is exposed to stimuli that cause binocular rivalry in human cases, e.g., a face to one eye and a starburst to the other. The “report” made by monkey’s button pressing suggests it does experience binocular rivalry.

The activity of the monkey’s neurons is also monitored. At the training stage, the neurons responding selectively to one kind of stimuli (e.g. faces vs. sunbursts) are pinned down. Then, in the binocular rivalry stage, the neurons which were previously found to selectively respond to each stimulus are monitored. Combining the data obtained from monitoring and the data from the “reports,” it was found which neurons were active during the “reports.”

They found that cells in the primary visual cortex do not correlate well to the monkey’s percept. Rather, further up in the processings of the brain, the activities of many cells in the inferior temporal cortex are strongly correlated with percepts. That is, some cells in inferior temporal cortex fire only when the monkey sees faces, and others fire only when it seems sunbursts. The results naturally suggest that neural correlate of visual consciousness includes the inferior temporal cortex.[8]

The reasoning here can be summarized as follows. Phenomenological description, or “reports” in this case, suggests that the monkey experiences certain stimulus. Independently from this report, representational properties of the brain are fixed. Finally, by combining the report and the monitored neural activity, the subset of neurons involved in representing a certain stimulus are selected as the neural correlates of the visual consciousness of the stimulus. In this case, phenomenological description and neural findings support each other. In other words, mutual constraint is satisfied. The phenomenological description, in this case the “report” of the monkey, suggests that it is experiencing a certain type of stimulus. This in turn suggests that representation of the stimulus in the brain is involved, which is ascertained by neural findings. Thus, the key trick here is that the experiment concerns representational aspects of the experience. This is directly related to representational properties of the brain, which can be at least partly independently investigated from reports. This mutual constraint applies to the

[8] Although activity in the primary visual cortex does not correlate with whether a given stimulus is a face or a sunburst, it is possible that it is a neural correlate of other aspects of visual consciousness. The predictive processing framework may support this line of reasoning, as its hierarchical structure suggests different cortices are responsible for different aspects of perception (see the section on Hierarchy in the next chapter.)

five Ps because the five Ps also concern *representational* aspects of phenomenology.

The mutual constraint between the phenomenological and the neural becomes much more difficult when it comes to other aspects of phenomenology that are not standardly treated as representational. For example, consider pain. The neural correlates of depressive mood may be found (perhaps in the anterior cingulate cortex), but this does not straightforwardly *explain* the phenomenology of pain because it does not account for why activation of a certain kind gives rise to the particular phenomenology of pain rather than the phenomenology of pleasure.

This mutual constraint plays an important role in the description problem for the five Ps. A good example comes from Poverty. Suppose it is claimed that you experience a richly detailed scene throughout your visual field. Then, you should experience a coffee cup placed at the far left or right of your desk with the same quality of vision as the computer screen in front you. That I experience the coffee cup in a fine-grained way places some constraint on computation in the brain. It mandates the existence, in the brain, of a fine-grained representation of the shape and color of the coffee cup. If there is no such representation, then the description is wrong; it is a false belief about conscious perception[9]. Put more generally, phenomenological description has to be compatible with the mechanisms in the brain described by predictive processing.[10]

This proposal is similar to Varela's (1996) notion of *neurophenomenology*, according to which accounts of phenomenology and accounts in cognitive science should be mutually constrained. Both my proposal of mutual constraint and Varela's view place importance on phenomenology but there are some important differences. First of all, Varela takes a Phenomenological approach (Capital P), which refers to the specific philosophical school originated by Edmund Husserl and developed by his

---

[9] If it turns out that it is a false belief, a fuller account of the phenomenon requires accounting for why the false belief comes to be held by the subject.

[10] Things might be able to work in a more bottom-up way; predictive processing suggests how perceptual phenomenology should be given representations in the brain. This is in a similar spirit to the bottom-up approach for theories of consciousness sketched in Hohwy (Hohwy, 2015), which recommends starting from the theory of whole brain functioning (predictive processing) and then looking for theories of consciousness that are compatible with the theory of the brain functioning.

followers. Varela adopts specific methodology in describing phenomenology (Varela, 1996). I am skeptical about its usefulness. For example, in the neurophenomenological tradition, it is asserted that we can explore possible forms of experience by imagination but it is doubtful that we can. For one thing, there is subjective variability in imagination; an experience imaginable to one might not be imaginable to others. It is also not easy to decide whose imagination should be taken as standard. Furthermore, the existence of bizarre experiences we could never imagine has been shown in pathological studies. For example, it is well known that selective disruption in the area MT cause akinetopsia (motion blindness) (Zeki, 1991). In a typical case of akinetopsia, even though the patient can experience static objects and their properties, motion is experienced like a sequence of snapshots. Odd experiences like this are unimaginable from the viewpoint of a healthy subject. Thus, I am sympathetic to Dennett's (1991, p. 44) criticism:

> Like other attempts to strip away interpretation and reveal the basic facts of consciousness to rigorous observation, such as the Impressionistic movements in the arts [sic] and the Introspectionist psychologists of Wundt, Titchener and others, Phenomenology has failed to find a single settled method that everyone could agree upon.

Secondly, as we have seen, my proposal is limited to representational aspects whereas Varela seems to try to apply it to every kind of experience. I fleshed out the mutual constraint idea for representational aspects and noted that we should not expect similar kinds of constraint in nonrepresentational aspects.

Thirdly and relatedly, Varela believes neurophenomenology and the mutual constraint idea can close the "gap" between the physical and the mental and thus constitute at least a partial solution to the hard problem. I do not think that is the case. When it comes to representational aspects, accounts of representational properties of the brain are necessary (though possibly not sufficient) to explain how our conscious perception represents the world. However, when it comes to apparently nonrepresentational aspects, it remains a mystery why the activity of certain regions gives rise to certain kinds of phenomenology. Therefore, while Varela's neurophenomenology and my approach are methodologically similar, their scope and metaphysical commitments differ.

What I think is a minimal metaphysical commitment is the principle Chalmers (1995) calls the principle of *structural coherence*. This principle presupposes isomorphism between the structure of awareness and structures of consciousness. Awareness here is defined as a complete functional term. Chalmers says, "[T]he contents of awareness are to be understood as those information contents that are accessible to central systems, and brought to bear in a widespread way in the control of behavior (Chalmers, 1995, p. 212)." This principle can work as a springboard from neurocomputational accounts to conscious experience. If we can pin down informational processing regarding the identity of a person in the brain, for example, then we can assume the subject is having the relevant experience. In this way, with the principle of structural coherence, predictive processing accounts provide an explanation for relevant conscious experiences.

Apart from "mutual constraint" discussed above, there are other benefits to applying the predictive processing framework to solve the three problems other than the description problem. Regarding the inverse problem, the answer is straightforward. As we will see in detail in the next chapter, the theory of predictive processing was designed to solve that problem.

Regarding the perceptuality problem, predictive processing makes a unique contribution. The perceptuality problem thus far received little attention outside philosophy and because of that, it has been discussed in mostly folk-psychological terms. Empirical evidence is usually considered, but theoretical perspectives have not often been taken in (except for the recent surge of works incorporating predictive processing).

However, there are reasons to consider more theoretical perspectives. The five Ps are undoubtedly part of our daily experiences but they are subtle aspects of perception that show different characteristics from more conventional aspects such as color or shape. To explain those subtle aspects, folk-psychology might not be an apt tool. Also, as it turns out, representations involving the five Ps show intermediate, borderline characteristics of folk-psychological categories. Predictive processing enables a principled explanation of such representations, as it suggests that the distinction between percepts and concepts be smeared (more on this next chapter). In a nutshell, by relying on predictive processing, we can give an account of each phenomenon that is free from the limitations of folk psychology.[11]

Regarding the reason problem, the fundamentality of predictive processing and the free energy principle grounds the reason why we have particular phenomenology. As I will introduce shortly, predictive processing is a special case of the free energy principle, which applies to any organisms that are in equilibrium with their environment. By looking at the free energy principle and how predictive processing falls out of it, we can see that predictive processing is equipped with particular elements. I will use these elements of predictive processing to explain the five Ps. In this way, we can gain insights into why we have perceptual phenomenology of the particular structure.

I have been stressing the benefits of explaining conscious perception using an empirical theory of mind, but this is not to say philosophical analysis or phenomenological observation is unnecessary. On the contrary, they play an important role in giving accounts. Without phenomenological observation, we cannot even find a phenomenon to explain. Careful conceptual clarification is also necessary to give a good account of a phenomenon, clearing up the messy ground of current debates. Therefore, conceptual ground cleaning is also one of the central tasks of my thesis. Armchair analysis should go hand in hand with empirical theories and evidence in building a theory.

So far, I have discussed how an account of the five Ps is benefited by the framework of predictive processing. However, these five Ps are also a challenge for predictive processing. Predictive processing is meant to be the theory of the whole brain. Therefore, we can say that if predictive processing accounts of the five Ps fail, then predictive processing is false. On the other hand, if predictive processing accounts succeed, they add to the explanatory power of the theory and the theory becomes more credible. Predictive processing is still a new theory and it is an ambitious one. Being able to explain such high-level phenomena is a necessary requirement for the theory.

---

[11] Paul Churchland (1981, 1989) has argued for the position called eliminativism. Churchland argues that folk-psychology is not a good empirical theory of the mind and should be replaced by a future complete neuroscientific theory. This thesis does not go so far as Churchland's. Admittedly, predictive processing is one promising theory to be the complete theory and it might replace folk-psychological theory. However, it is not conclusive yet that folk-psychology is a successful empirical theory.

It is sometimes claimed that predictive processing adds nothing to local, domain-specific theories such as existing motor control theories for action, because local theories can do as good an explanatory job as predictive processing. The worry behind this criticism also concerns the alleged fundamentality of the theory; if predictive processing is simply compatible with existing accounts, one might want to avoid huge theoretical commitment of predictive processing. However, this criticism focuses only on the mechanistic how questions. On the one hand, there is the virtue of unification itself (Kitcher, 1989). The most influential cases of unification are Newton's theory of physics and Maxwell's theory of electromagnetism. The former unified terrestrial and celestial motion, which were thought to be subject to different laws; the latter unified electricity and magnetism, also thought to be separate phenomena subject to different laws (Woodward, forthcoming). According to Kitcher (1989), the virtue of unification consists in reducing the number of primitive facts and patterns of reasoning from which we can derive descriptions of different phenomena. Predictive processing precisely does this, and my thesis reproduces this pattern. Predictive processing accounts of the five Ps are written in the limited terms of predictive processing, thus predictive processing accounts will show deep similarity between the phenomena. On the other hand, most problems I have introduced so far—the problem of perceptuality, the problem of description, and the problem of reason—cannot be addressed by local theories of cognition. We have seen how predictive processing can contribute to the solution of each problem and the contribution is made possible because it is a general, fundamental theory of cognition. Although, I assume the predictive processing framework in this thesis, the success with which it accounts for and provides solutions to each of the problems of interest certainly adds to the theoretical credibility of the predictive processing framework.

2. *Preparing the Predictive Processing Toolbox*

In this chapter, I introduce the predictive processing framework and the elements of it that I will use to account for the five Ps presented in Chapter 1. First, we will look at the free energy principle. Then, we will see how free energy principle entails predictive processing.

2.1. *The free energy principle and predictive processing*

The predictive processing framework I work with mainly concerns how the brain comes to represent the world, but the free energy principle explains, from a wider perspective of biological systems, why the brain engages in such inferences at all. Bayesian inference is closely related to the free energy principle (see Friston (2011, p. 98)), and predictive processing, an approximation of Bayesian inference, is also entailed with some assumptions under free energy principle. Therefore, looking at the free energy principle helps us to understand how and why the brain does predictive processing. In this section and the next, we see what the free energy principle is and how it entails predictive processing under certain conditions. Then, I will introduce the core features of predictive processing in Section 2, which will help account for the five Ps.

The free energy principle is the fundamental principle that explains why biological systems are the way they are. The free energy principle itself is fairly simple, stating that, "any self organizing system that is at equilibrium with its environment must minimize its free energy" (Friston, 2010, p. 1).

The free energy principle begins from the tautology that living organisms are, at least to some extent, successful in survival. For survival, biological systems need to maintain their homeostasis in the face of changes in the environment. Successful organisms are in limited sets of internal states (i.e. physiological and sensory states). For example, a fish tends to be in water and a lion tends to be on the savanna. And by virtue of being in limited kinds of environmental states, they are also in limited kinds of physiological and sensory states (such as sensory registration from external stimuli or body temperature that suits the environment.) Thus, a probabilistic distribution of physiological and sensory states of an organism can be described. A lion is found more on the savanna rather than in the sea, a fish is found more in water than up on a hill.

Moreover, for a fish to be on land is not good for its survival. Thus, a fish needs to avoid being on land. Put more generally, an organism has to avoid finding itself in an internal state that has low survival probability given what animal it is. This is equivalent to say that an organism has to avoid surprise[12]. Thus, success in survival is to maintain internal states within certain bounds in the longer term, and so the biological system has to avoid surprise in the short term. The problem here is that a system does not have direct *access* to surprise, and therefore cannot respond to it. The system has *access* only to its internal states and sensory states; even its own bodily states are not directly accessible—they have to be inferred from somatosensory states. On top of this indirectness, the computational burden is immense: calculation of surprise requires consideration of every possible state at any given time. This is an unattainable task for biological systems such as the humans, brain.

This is where free energy comes into the picture. Free energy is a mathematical function of sensory states and recognition density. Recognition density is a probability distribution of represented causes of sensory inputs in the environment encoded by its own neural state. Free energy is accessible to the system because sensory states and neural states are within the system. Moreover, free energy sets an upper bound on surprise (That is, free energy ≥ surprise.) Therefore, if you can minimize free energy, you can thereby minimize surprise implicitly.

Free energy can be minimized by changing two variables in the free energy formulation: changing sensory states and changing the recognition density. One way is by changing sensory samples so that they conform to predictions. The other way is by changing the recognition density in order to change conditional expectations about sensory samples. These two ways roughly correspond to action and prediction respectively, though the distinction is somewhat more nuanced. I will come back to this point shortly.

Friston (2010) discusses the characteristics of free energy illuminated by different mathematical formulations of free energy and shows the relationship between the free energy principle and Bayesian perspectives, including predictive processing.

One way to formulate free energy mathematically is to see free energy as energy minus entropy. In this formulation, you can see that free energy rests on the generative

---

[12] Formally, surprise is a negative log probability of an event. Let surprise be h, an event be r, and the probability of the event $P(r)$. Then, $h(P(r)) = -\log P(r)$.

model, which predicts sensations given a model of causes in the world and their probability. “Energy” (different from free energy) is the surprise about the joint occurrence of sensations and their represented causes and “entropy” is entropy of recognition density. Energy represents how likely or unlikely sensations are given a representation of their causes and their expected probability (in Bayesian terms, it is a likelihood function). If a rabbit in front of a perceiver is the cause of a sensation, it should cause a characteristic pattern of sensation in the perceiver.

The generative model can be decomposed into two parts: the prior probability of the model and the likelihood of a particular sensation given the model. Here, prior possibility corresponds to recognition density and is thus related to entropy. Likelihood corresponds to joint probability and is thus related to energy. Therefore, if the sensations are well predicted by the generative model, “energy” becomes lower and “entropy” becomes higher and thus free energy is thereby minimized.

Another way is to see free energy as surprise plus divergence. Divergence is always non-negative (0 or positive), thus free energy sets an upper bound on surprise. Divergence is the difference between the recognition density and the conditional density of the causes of a sensation given the sensory signals. The conditional density represents the best possible guess about the true cause of experienced sensory stimulation. Thus, free energy can be minimized by changing recognition density in order that the density approximates conditional density. This rendition shows that if the recognition density becomes the conditional density, the brain represents the best possible causes given sensory stimulation and so minimizes free energy.

The third and the last formulation regards free energy as complexity minus accuracy. Complexity is the difference between the recognition density and the prior density on causes but what is important in this rendition is “accuracy.” Accuracy is the surprise about predicted sensations under the recognition density. Thus, minimizing free energy by changing sensory data without changing the recognition density must increase the accuracy of an agent’s predictions. In other words, when the agent selectively samples the sensory inputs that its generative models predict, free energy is minimized.

These three ways of formulation already imply that the brain approximates Bayesian inference. In the first formulation it was proposed that if generative models of the brain can successfully predict sensory states, free energy is minimized. The model that gives the highest posterior probability (which is the best guess given sensory states) minimizes free energy and is thus selected. This is Bayesian inference, and follows this formula:

$$P(m|s) \propto P(s|m)P(m)$$

Where m symbolizes a model of the causes, s symbolizes sensory states, and $P(m|s)$: posterior probability, $P(s|m)$: likelihood function, and $P(m)$: prior probability.

Below is a summary of the free energy principle:

- For survival in the long run, biological systems minimize surprise. That is, a biological system tries to stay within a limited set of internal states.

- Biological systems cannot assess and minimize surprise directly. Instead, they minimize free energy (where free energy ≥ surprise), which the systems have direct access to.

- Free energy is a function of recognition density and sensory inputs. Thus it can be minimized in one of two ways: by changing recognition density (which roughly corresponds to perception) or by changing sensory inputs (which roughly corresponds to action).

- When recognition density is changed to minimize free energy, it has to be changed so as to represent the most likely causes of the experienced sensations in order to minimize divergence.

- When sensory inputs are changed to minimize free energy, sensory inputs have to be changed so as to fit predictions to maximize accuracy.

- Free energy minimization implies Bayesian inference. Selecting generative models that best explain sensory states just is performing approximate Bayesian inference.

2.2. *Introducing Predictive Processing*

As we have observed, approximate Bayesian inference is implied by the free energy principle. In other words, performing Bayesian inferences is one of the optimal ways for an organism to survive. By performing approximations of Bayes rule, the brain is able to represent the causes of sensation in the world.

I proposed above that minimization of free energy is accomplished in one of the two ways, corresponding to perception and action. These two aspects will be fleshed out in predictive processing terms in Section 2.4. In what follows, we see what predictive processing is and also prepare a predictive processing "toolbox", which in later chapters will be used to explain the five Ps.

The idea of perception as Bayesian inference has important implications for what perception is and how conscious percepts are formed. Predictive processing is essentially a Kantian, constructivist theory, in that it assigns a large role to the subjective take on perception (i.e. priors in Bayesian inference) and it involves a generative model that explains sensory stimulation (Fazelpour & Thompson, 2015; Schlicht & Newen, 2015). What you perceive depends on the "subjective take" of your brain about the world, rather than exclusively on the sensory stimulation you as a perceiver receive from the world.[13] In this sense, predictive processing makes a stark contrast to more empiricist views of perception. In more classic theories of perception (such as Marr (1982)), perception is conceived in a more bottom-up, empiricist fashion. Such theories claim that we receive sensory information, and representations are incrementally built upon that information. But in the predictive processing framework, there is an emphasis on "prior ideas" that the brain has about the causes of sensory inputs. However predictive processing does not go so far as idealism since the brain still has to explain or predict actual incoming sensory stimulation. Thus percepts need not be wholly determined by sensory stimulation; as long as the prediction can explain sensory stimulation there is some leeway for the brain to represent the world differently. This characteristic of predictive processing plays a pivotal role in Chapters 2 and 3.

[13] Surely, by saying "subjective take" here, I don't mean anything like conscious understanding. I rather mean that the representation is formed based on priors and sensory samples, and the brain does not have direct access to external objects and their properties. See also my terminological caveat below.

2.2.1. Prediction error minimization

Predictive processing involves a particular way that a brain approximates Bayesian inference. Strict Bayesian inference is computationally very costly, so the brain is implementing some approximation of Bayesian inference. One mechanism is to limit the form of probability density that is used to explain sensory samples. If the brain only uses *Gaussians* (or normal densities), it becomes computationally cheaper. A probabilistic density is usually encoded by its *sufficient statistics*. Sufficient statistics can be considered parameters specific to a form of probability densities, which are sufficient to fix those probability densities. In the case of Gaussians, sufficient statistics are only the conditional mean and variance. That is, if the mean and the variance are determined, there is only one possible Gaussian. Conversely, other forms of probabilistic density can require many more parameters to yield sufficient statistics. Limiting the kind of probabilistic density is important for two aspects. Firstly, as already mentioned, the more parameters are required, the more computational power is required. At the same time, the fewer sufficient statistics are required, the stronger the assumption on the form of density becomes, which then limits explanatory power. However, the simplest density, Gaussian, is the most frequently assumed (called the Laplace assumption or approximation). Under this assumption, a Bayesian inference becomes just a matter of minimizing the difference between the model's predictions about sensory samples and the sensations (using only means and variance). In this way, minimizing free energy collapses into minimizing prediction errors in the long run, and this forms the core of predictive processing.[14] There is now much evidence that the brain does this kind of processing. For example, it has been argued that many features of early visual responses are explained by this framework (Rao & Ballard, 1999; Sherman, Seth, & Kanai, 2016) and that the framework provides a plausible account of repetition suppression and mismatch responses in electrophysiology (Auksztulewicz & Friston, 2016; Rao & Ballard, 1999; Sherman et al., 2016; Stefanics, Kremláček, & Czigler, 2014).

[14] The brain also uses other techniques to reduce computational burden. It is thought that the brain implements variational Bayes to avoid calculating high-dimensional integrals.

Prediction error minimization is the first component of predictive processing that goes into the five P toolbox, and is defined as follows:

> *Prediction Error Minimization*
> The brain predicts sensory inputs given a generative model. The difference between predicted sensory inputs and actual sensory inputs is prediction error. The brain needs to minimize prediction error over the long term.

A terminological caveat needs to be made here. The predictive processing framework, or Bayesian approaches to cognition in general, are full of terms that are used to refer to high-level cognitive phenomena, such as "inference," "prediction," "expectation," and "belief." But the use of these terms in the Bayesian framework does not refer to the conscious cognitive phenomena that are usually referred to by those terms. On the contrary, they refer to unconscious computational processes in the brain. It is not that the terms are used in an equivocal way, it is just that the terms are used in a special, technical sense. Their use in this way is widely accepted in the machine learning literature, and the processes referred to have some similarity to their conscious variants. For example, when I say "Bayesian inference", this only means the brain implements some computational processes whose behavior is described as following (or approximating) Bayes' rule. When I say the brain or the model "makes a prediction", that doesn't mean making a conscious prediction. This only means that the generative models pass a message top-down, which encodes sufficient statistics of some density. Hereafter, when I talk about conscious phenomena, I will explicitly note this.

2.2.2. Hierarchy

Flexible learning by performing approximate Bayesian inference is achieved in a hierarchical setting (Mathys et al., 2011). Hierarchy also provides an answer to an objection to Bayesian approaches for perception and cognition (Friston, 2011). The objection asks how prior beliefs are obtained. Friston writes,

> [C]rucially, because empirical priors are linked hierarchically, they are informed by sensory data, enabling the brain to optimize its prior expectations online. This optimization makes every level in the hierarchy

> accountable to the others, furnishing an internally consistent representation of sensory causes at multiple levels of description. (Friston, 2011, p. 99)

The hierarchical structure splits prediction error minimization into multiple levels. Each level explains/predicts the bottom-up signals coming from the level below, and whatever cannot be predicted at that level (prediction errors), are sent to the next level above up. The prediction errors are the bottom-up signal for a higher level. The bottom up signals for the lowest level are the sensory signals from the world, by which the systems is connected with the world. The hierarchy is regimented by a timescale of causal regularity that a level concerns; the higher in the hierarchy, the longer the timescale of causal regularity the level deals with. Moreover, there is a connection between the timescale of regularity that is represented and the degree of detail the contents have. The faster timescale regularity is good at capturing at details whereas the slower timescale regularity is suitable for more abstract, general content.

This difference in timescales can be tied with (non)perspectivalness (or the spectrum between variant and invariant representation). Perspectival information fluctuates as the relationship between the perceiver and the objects changes, but nonperspectival information is by definition something that is immune to such changes. This in turn means that the contents of representation at higher levels are more abstract and thus more concept-like, and, in contrast the contents of representations at lower levels are more detailed and percept-like. Thus the multi-level hierarchical nature of the system suggests that the boundary between perception and cognition is blurred despite its folk-psychological clarity. As I will argue shortly, however, *percepts* and *concepts* are matters of degree, but *perception* and *cognition* are not.

One caveat about hierarchy is that a higher level is not "about" the level below it, and thus the hierarchy itself does not involve meta-representations. Rather a representation at each level represents different aspects of sensory causes regimented according to timescale, whereas a complete percept emerges as the sum total of representations across the hierarchy.

We will see shortly that *Hierarchy* in tandem with *Precision Expectation* (explained next) plays a constitutive role in explaining various mental phenomena. Thus, Hierarchy is an important component of the predictive processing toolbox. Here is the definition of *Hierarchy* **(**Figure 2**)**.

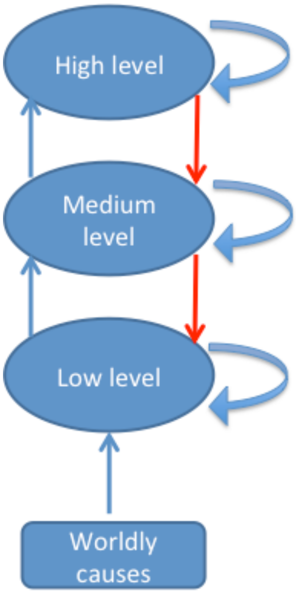


Figure 3 The hierarchical structure of predictive processing system

Hierarchy

The biological prediction error minimization system has a hierarchical structure. Each level predicts bottom-up signals coming from the level below. Prediction errors, what cannot be predicted at that level, are sent to the level above, and those errors form the bottom-up signal for the next level. The hierarchy is regimented by timescales of causal regularity; the higher in the hierarchy, the longer the timescale of causal regularity the level deals with. Perception is determined conjointly by multiple levels of representations in the hierarchy.

2.2.3. Precision Expectation

As I discussed above, sensory samples are not obtained all in one go. Thus the brain needs to keep updating predictions made by the generative model as it is given new sensory samples. The problem is how much of the prediction error should be reflected in the model—how it should be weighted. This is not easy to determine because it requires assessment of how much the system has already learned. When the system has just started learning, the model should weight prediction error a lot, but if the system is deemed sufficiently knowledgeable about the cause in question, prediction error should tend be dismissed—when the system is already confident, prediction errors may occur because the sensory samples are statistical outliers. Thus, the learning rate of prediction error should be set with reference to the system's take on how much it has already learned. In a predictive processing setting, the learning rate is set in accordance with the precision of priors. The higher the certainty about the priors is, the less the brain updates its current model; the more the current sensory samples or prediction error are well predicted by a model (i.e. its precision is high), the more the brain updates its current model.

In addition to reliance on the prior precision, Bayesian inference also must rely on the likelihood precision, which reflects the precision or variance in the sensory input itself. High likelihood imprecision should decrease the perceptual learning rate. The brain can incorporate contextual influences on this learning rate setting, or precision estimation, by making use of the overall hierarchical setting. The sensory signals the brain receives contain noise; they do not always reflect causes in the world and can instead be affected by other intervening factors, be the external (e.g. lighting conditions) or internal (e.g. conditions of sensory organs). This is crucial because the level of precision must be assessed relative to a particular context. That is, the learning rate should be set in accordance with the brain's estimation about the context regarding whether current prediction error is trustworthy. As I introduced above, higher levels of the hierarchy concern longer timescales. Therefore, a higher level can provide a lower-level contextual information not only about content but also about the precision of prediction error at that level (e.g. "This is a poorly-lit room, so I cannot trust the apparent surface color of objects in this room"). Therefore, precision expectations under a hierarchical setting enable flexible learning about the changing world, and can be defined as follows:

> *Precision Expectation*
> The system estimates how precise prediction error is to decide how much prediction error should be taken into consideration in the model. Contextual information about precision is taken in by higher levels of the model.

Recently, it has been argued that *attention* is tightly linked to the optimization of precision estimation (or simply precision optimization) in the hierarchical system, or that precision expectation simply *is* attention (Feldman & Friston, 2010; Hohwy, 2012). Attention is a catch-all term that encompasses notoriously various phenomena (Mole, 2011). Here, I focus on only two such phenomena. When you hear a sudden loud sound from the street, it instantly captures your attention. When attention is grabbed by an external event, this kind of operation of attention is called *exogenous attention*. This can be explained in terms of precision optimization as follows. Stimuli that grab attention tend to have a better signal to noise ratio, and it is reasonable to expect such signals to be precise. Thus it is reasonable to expect any ensuing prediction error to be precise and worthy of high weighting in perceptual inference.

Attention can also be driven more intentionally. You might be told at a festival, "The fireworks will be seen in that direction", causing you to attend to that area of the night sky. This is an example of *endogenous attention*. By virtue of the prior verbally obtained information, the stimuli subsequently occurring in that area are expected to have good signal to noise ratio, and its prediction error will be estimated to have higher precision. The relationship between attention and precision optimization will be discussed further in Chapter 3.

The estimation of precision and hierarchy is considered related to mental disorders. For example, the autism spectrum and many delusions may be related to failures in precision estimation (Corlett, Frith, & Fletcher, 2009; Gerrans, 2015a, 2015b; Hohwy, 2013). According to Hohwy, delusions arise because of an underestimation of the precision of prediction errors. This underestimation leads to a failure of taking in prediction errors. That is, even though there is sensory evidence that goes against the content of the delusion, it is disregarded because of the low precision expectation. On the other hand, aberrantly high precision estimation is related to autism spectrum (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; Van de Cruys et al., 2014). Autistic people are hypothesized to constantly revise their models in the face of prediction errors because of the high precision settings. This leads to difficulties in disregarding prediction errors, as would happen when taking in context and larger perspective. This explains not only their social dysfunction but also their sensory profiles. I will return to the topic of delusion in Chapter 6.

2.2.4. Active inference

We have already seen that there are two ways in which free energy can be minimized, corresponding roughly to action and perception. I have also argued that minimization of free energy by action occurs when sensory samples are changed so as to correspond to expectation. In the cases discussed so far, the model has been modified to fit the sensory samples. I call this *perceptual inference*. Sensory samples can also be changed to fit the models through minimization by action, which I call *active inference*. In other words, perceptual inference and active inference have opposite directions of fit—perceptual inference has a direction of fit of the mind to the world, whereas active inference has a direction of fit of the world to the mind.

Active inference works as follows. The hypothesis that currently has the highest probability is used to make a counterfactual prediction about possible sensory stimulation dependent on action. Then, when the action is actually performed, the hypothesis is confirmed or refuted. For example, when you are facing a car from an angle, your brain has some idea about how the car would look like from other angles. When you then move around the car, the predictions are confirmed (in this case).

By performing active inferences, you can efficiently reduce uncertainty about the model. If you only rely on perception, reducing uncertainty takes time because of the limitation of available sensory stimulation as well as fluctuations in noise level. By performing actions, however, the system can rapidly gain new sensory stimulation and (dis)confirm hypotheses quickly. This is particularly useful when the winning (correct) model does not have a higher initial probability than competitors.

I suggested above that perceptual inference and active inference roughly correspond to the distinction between perception and action, but the story is not so simple. It turns out that active inference plays an important role in conscious perception. Active inference includes unconscious behavior such as saccades, which are modeled as an instance of active inference (Friston, Adams, et al., 2012). In this framework, saccades are conceived of as experiments testing hypotheses. Saccades are an integral part of perception. Thus, conscious percepts are the joint product of perceptual inference and active inference[15]. Active Inference also makes it into the toolbox of predictive processing, and is briefly defined below:

[15] This conception of perception essentially involving action echoes Alva Noe's sensorimotor theory of perception. The connection between these theories is further explored in Chapter 3.

Active Inference

The brain performs action based on counterfactual predictions derived from the model that currently has the highest probability. By performing active inference, the system can quickly reduce uncertainty about the model.

2.3. *Perception and cognition under predictive processing*

In the previous sections, I introduced the basic components of predictive processing. However, to illuminate the core topic of this thesis, we need to see the implications of the theory for *perception*.

In the predictive processing framework, the content of conscious perception is *determined* by the content of the model that, at a given moment, has the highest posterior probability. For example, a subjective conscious percept during binocular rivalry that does not correspond to environmental stimuli is explained in terms of the model that has the highest posterior probability (Hohwy, Roepstorff, & Friston, 2008). Content here is phenomenal representational content defined as follows.

It is common to think that perceptual experience represents the world as being in a certain way by virtue of its phenomenology (Davies, 1992; Tye, 1995): the world phenomenologically seems to a subject to be a certain way. However, phenomenological seeming is not necessarily correct. For example, it might seem to you that there is a tomato in front of you where there isn't one. In such cases, the experience is non-veridical. Mostly, we believe our experiences are veridical, such as when it seems to you that there is a tomato in front of you and there actually is a tomato. I call this aspect of perceptual experience, assessable in terms of veridicality, *phenomenal representational content* or *phenomenal content* (more on this in Chapters 4 and 5). It might not be that all the contents of the model that best explains sensory stimulation are consciously experienced. Some contents are too abstract to be experienced (e.g. the earth is round), and some contents are too detailed to be explained (e.g. content of retinal representations). However, some contents of the models are undoubtedly experienced. Throughout this thesis, I assume that not only representations about shape and color, but other kinds of attributes such as object category and personal identity are perceptually experienced.

Another important consequence of predictive processing concerns the relationship between perception and cognition. In the above discussion about perceptual *Hierarchy*, I mentioned that the boundary between percepts and concepts gets blurred because of the multi-level nature of the hierarchy. This hierarchical setting of prediction error minimization and precision expectation naturally suggests the top-down cognitive involvement of perception. As was explained in the section on Precision Expectation, when uncertainty is higher, prediction error is not taken into a model very much, and thus the priors mostly determine the posteriors in those cases. The priors are furnished by models at a higher level. Thus top-down, more cognitive models take part in forming conscious percepts. Clark nicely encapsulates this point:

> To perceive the world just is to use what you know to explain away the sensory signal across multiple spatial and temporal scales. The process of perception is thus inseparable from rational (broadly Bayesian) processes of belief fixation, and context (top-down) effects are felt at every intermediate level of processing. (Clark, 2013, p. 10)

This aspect of predictive processing is not an independent component of prediction error minimization; it is implied by *Precision Expectation* and *Hierarchy*, and some weighting of top-down priors is necessary in Bayesian inference. Since it will play a key role in some chapters (mainly Chapters 5, 6, and 7), I will add cognitive penetration to the toolbox:

> *Cognitive penetration*
> High-level representations influence the posterior probability in perceptual inference; this influence increases when the precision (or expected precision) of the bottom-up prediction decreases.

The top-down involvement of cognitive penetration in perception reveals how different levels join to form the whole percept and this may warrant the claim that the boundary between perception and cognition is fuzzy. However, this might not yet show that there is no boundary between *perception* and *cognition* (as Clark (2013, p. 10) suggests). More paradigmatically cognitive cases such as conscious thinking or conscious imagining are still left out of the picture. Surprisingly, despite the ubiquity of cognitive metaphors in predictive processing and the Bayesian models of decision-making, thinking or imagining is yet to be accounted for within the predictive processing framework. A full explanation of them is beyond the reach of this thesis, however, we may spell out what they are not. When we are entertaining conscious thoughts or imagery, we are not perceiving. In other words, we are ignoring or segregating out stimulation at the sensory organs and are not primarily in the business of representing the immediate world surrounding us when we are engaged in thought. This characterization of conscious cognitive states echoes Clark's statement about perception quoted above, "To perceive the world just is to use what you know to explain away the sensory signal across multiple spatial and temporal scales." Thus perception can be a matter of explaining away current sensory stimulation and cognitive states are something that are not in the business of explaining current sensory stimulation. (Notice that this definition is not entirely satisfactory, since dreaming and imagery, for example, are also not a matter of explaining away current sensory input; however, it will do as a working definition). Hence, we get the following definition of perception, which I will use in this thesis:

> *Perception in predictive processing*
> The content of perception is the contents of models at all levels that are used to explain away current sensory input.

This suggests that the boundary between percepts and concepts is fuzzy, but the boundary between perception and cognition is not. Perception can contain higher-level concept-like states but as long as all the levels of the hierarchy are used to explain away current sensory stimulation, it is nevertheless clearly a case of perception.

How can illusion and hallucination be explained under this notion of Perception? The content of perception is the content of the models that explain sensory samples the best. However, sometimes the best guess from the sensory samples does not lead to veridical perception.

This happens in cases of illusory perception, where sensory stimulation is similar to that which would be produced by another cause, giving rise to non-veridical perceptions. Take the Müller-Lyer illusion as an example (Figure 4). One standard account of the Müller-Lyer illusion is to regard it as a misleading version of size constancy (Hohwy, 2013). The wings of the arrows work as the cues signifying that the lines are placed at different depths. In the natural world, such cues are right and helpful most of the time. In the natural world, sensory stimulation obtained from inward pointing wings is obtained from an object placed in front, and sensory stimulation obtained from outward pointing wings is obtained from an object at depth. In a 2D image, the brain uses the depth cues provided by these wings to counterbalance the objects, making the object in the back look larger it is and the one in the front look smaller. Thus in **Figure 5** below, the length of WALL A is longer than that of WALL B, but if the image is interpreted as being 3D, they look almost at the same length. Similarly, in the Müller-Lyer illusion, the 2D length of both arrows is the same, which results in percept that an arrow with inward pointing wings looks shorter.

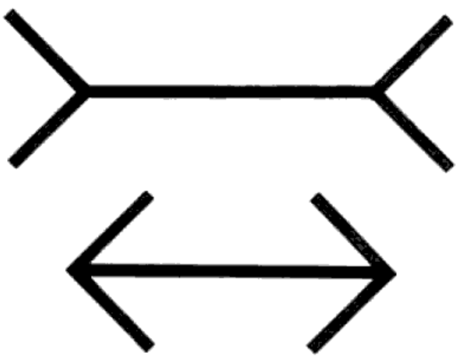


Figure 4 Müller-Lyer illusion

The point of this explanation is that *most of the time* when that kind of sensory stimulation is obtained, the brain's adjustment to make close objects smaller leads to veridical percepts. Over a long term average, therefore, this perceptual rule minimizes prediction errors, even if the rule generates illusions in rare cases.

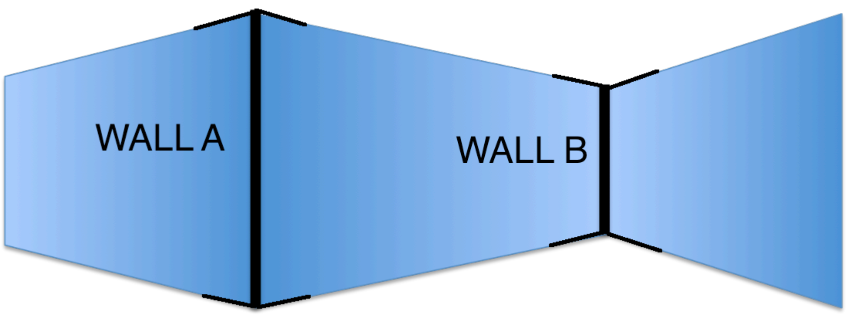


Figure 5 Size constancy

Importantly, prior conceptual knowledge that the lines in the Müller-Lyer illusion are of equal length doesn't change the percept, because for such cases, the precision of prediction error at low levels is high (there is no noise and size constancy normally leads to veridical percepts, cf. *Cognitive Penetration*). Veridical perception can therefore only be obtained when the subject, relying on the high-level belief that the lines are of equal length, intervenes in the environment by performing an *Active Inference*. To do this, you could take measurements of both lines or you can add two horizontal lines passing through the tops and bottoms of the vertical lines, thereby revealing that the lines are the same length. Such actions reveal that the vertical lines are of the same length, but both interventions change the original percept. To summarize, illusion (or hallucination) can take place when the brain receives sensory stimulation similar to that of veridical perception. This can be resolved by performing *Active Inference* based on veridical beliefs.

This definition might look revisionist, but it preserves a more traditional conception of perception. For example, it explains cognitive *im*penetrability. Cognitive penetrability introduced above does not imply that cognitive states such as conscious thoughts can directly alter perception. On the contrary, cognitive states shape perception only when 1) precision is estimated to be low, or 2) higher-level models can explain away prediction errors from the level below. In cases of illusion, the illusion persists despite conscious beliefs to the contrary. This is because the hypothesis pointing to the illusion is the one that has the highest posterior probability in explaining the sensory stimuli. There is therefore not much left to be explained by competing hypotheses, and so the contents of conscious thoughts are not called upon. I will return to this issue in Chapter 4.

There are some potential objections to my proposal. One possible objection points to the possibility that we can engage in conscious thinking to explain away sensory stimulation. If conscious thinking can explain sensory stimulation, the proposal fails for conscious perception. A potential case for this would be illusion. Faced with the Müller-Lyer illusion, you might persuade yourself that the top arrow is equal in length to the bottom arrow and try to explain away the *percept*. However, this is not a case of explaining away *sensory stimulation*. We can certainly consciously explain away conscious percept, but this is not the same as explaining away the sensory stimulation;

understanding the illusion does not stop it from intuitively seeming to be true. The conscious percept is the end product of the process of prediction error minimization, and explaining away sensory stimulation is an unconscious process to which a subject normally does not have conscious access (as I discussed in Section 2.2.1 (“Prediction error minimization”)).

Another potential objection might target my incorporation of higher-levels in perception. One might want to keep the realm of perception only at some intermediate-level (just as Jakendoff (1987) and Prinz (2012) do). However, keeping the realm of perception only at intermediate levels seems arbitrary from the predictive processing perspective. High-level models are not always recruited to explain away sensory stimulation, but when they are, they form a unified system with other levels. As we have seen in Section 2.2.2 ("Hierarchy”), will be explored in each chapter, the mechanistic rules that each level operates under are the same. A motivation for this objection may be to reserve high-level models for higher cognitions (such as thinking and reasoning). However, I make this distinction differently. High-level models may be used for higher cognition, but when they are, they are not used to explain away sensory stimulation. On top of that, my proposal can incorporate aspects of the traditional conception of perception (i.e. impenetrability).

We now have the five elements of predictive processing in our toolbox! I will use them in the following chapters to give accounts of the five Ps. Provided below is a table of this toolbox for reference (Table 2), showing the characters of each element of the predictive processing framework.

Table 2 The toolbox of predictive processing

Tools	Description
Prediction Error Minimization	The brain predicts sensory inputs given a generative model. The difference between predicted sensory inputs and actual sensory inputs is prediction error. The brain needs to minimize prediction error over the long term.
Precision Expectation	The system estimates how precise prediction error is to decide how much prediction error should be taken into consideration in the model. Contextual information about precision is taken in by higher levels of the model.
Hierarchy	The biological prediction error minimization system has a hierarchical structure. Each level predicts bottom-up signals coming from the level below. Prediction errors are sent to the level above. The hierarchy is regimented by timescales of causal regularity; the higher in the hierarchy, the longer the timescale of causal regularity the level deals with. Perception is determined conjointly by multiple levels of representations in the hierarchy.
Active Inference	The brain performs action based on counterfactual predictions derived from the model that currently has the highest probability. By performing active inference, the system can quickly reduce uncertainty about the model.
Cognitive Penetration	High-level representations influence the posterior probability in perceptual inference; this influence increases when the precision (or expected precision) of the bottom-up prediction decreases.

3. The First P: Presence

3.1. *Introduction*

The first P addressed in my thesis is *presence*. We looked at some examples of presence in Chapter 1, but let us consider another to introduce the topic here. Suppose you see a cat sitting behind a picket fence. You can only see its head, along with narrow slivers of the rest of the cat through the fence, but you can immediately recognize it as a cat—in this case, there is a sense in which you see the whole cat, not only the visible parts of the cat.

This is an example of *perceptual presence*. *Perceptual presence* refers to a range of phenomena highlighted by Alva Noë (2002, 2004, 2006), who used those phenomena as support for his own enactivist theory of perception. Perceptual presence is the target of this chapter. However, what is meant by perceptual presence and what problem (or problems) perceptual presence raises, is far from clear, mainly because the discussion is structured around individual case. Thus, I first offer a taxonomy of the phenomena involved to give a uniform account of perceptual presence. Then, I will give an account of perceptual presence based on predictive processing. I use the account to give a solution to two prominent problems associated with perceptual presence: the problem of representation and the problem of subjective veridicality.

The problem of representation concerns the forms of representations involved in perceptual presence (Nanay, 2010). The nature of perceptual presence as *perceptual* itself is sometimes cast in doubt. For example, Prinz (2012, p. 76) denies the perceptual nature of representation of the occluded aspects of the seen thing, relying on the intuition that it does not seem to him that he can *see* that part. It seems unlikely that the debate can be settled by debating conflicting intuitions, but it is undeniable that there are important differences between the perception of the seen front and that of the unseen back of an object. Only the experience of seen parts is vivid and salient. I will address the question of if perceptual presence really is perceptual by identifying the plausible mechanisms for the representation of the front of an object, the representation of the back of an object, and the representation as of the whole object. By doing so, we can advance the debate beyond conflicting intuitions.

I argue perceptual presence is a genuinely perceptual phenomenon. The representations or aspects of perceptual presence involved in perceptual presence are all explained as a result of *Prediction Error Minimization*, and all are employed to explain sensory stimulation.

The problem of subjective veridicality regards why certain representational contents feel real to the subject. Most instances of daily perception feel real but some do not. For example, when you play a first-person videogame from the 1990s, it does not feel quite real. The contents of the video game are experienced as only caricatures of actual things and creatures; they will never be conflated with reality. The current debate is about what actually gives rise to feelings of reality. I will argue against a recent account, also based on predictive processing, by Anil Seth (2014). My account gives a focus on coherence of the models the brain possesses.

Before moving on to taxonomy, I would like to touch on the implications the issue of presence has to other problems in philosophy of mind. It will have relevance to at least two important debates in philosophy of mind. One is the debate about the neural correlates of *perceptual* consciousness. If perceptual presence is genuinely perceptual, the neural correlates of it should also be included in the neural correlates of *perceptual* consciousness, and the general methodology concerning what counts as perceptual is also of high relevance. The other implicated debate regards the admissible contents in perception (Bayne, 2009; Siegel, 2012): the debate about what kinds of properties are represented in perception. In addition to this, the problem of subjective veridicality would contribute to our understanding of abnormal mental functioning, such as synesthesia and Charles-Bonnet syndrome. More generally, my account will offer a new, more integrated perspective on our experience of presence of objects in the world.

3.2. *Taxonomy of perceptual presence*

Many philosophers apart from Noë talk of perceptual presence, including Clark (2012) and Hohwy (2014), but it is undeniable that Noë set the tenor of the discussion. Noë talks of perceptual presence in many different places but he only focuses on giving descriptions of concrete examples rather than providing a clear unifying conception of the phenomenon. In this section, I will endeavor to give more systematic taxonomy of perceptual presence.

Let's begin with one typical example of perceptual presence, the one introduced at the beginning of this chapter.

> Occlusion
> A cat sits motionless on the far side of a picket fence. You have a sense of the presence of a cat even though, strictly speaking, you only see those parts of the cat that show through the fence.

In these *occlusion cases*, something hidden from sight is claimed to be perceived despite the absence of sensory stimulation from that object. In this example it looks as if only two kinds of representation are involved: the representation of visible parts and the representation of a whole cat ("you have a sense of the presence of a cat"). That is not correct. Even though this is often not well articulated in Noë's work and the broader debate about perceptual presence, there are actually three kinds of representation: the representation of non-occluded parts, the representation of occluded parts, and the representation of the whole object. I call these the tripartite representation of object.

For example, in one place, Noë writes:

> The visual experience of the tomato, when one takes it at face value, presents itself to one precisely as a visual experience *as of a whole tomato* [emphasis added]. (Noë, 2006, p. 413)

However, he also writes,

> How can it be true, as I think it is, that we are perceptually aware, when we look at a tomato, of *parts of the tomato* [emphasis added] which, strictly speaking, we do not perceive? This is the puzzle of perceptual presence. (Noë, 2006, p. 414)

An important question here is whether representing the front and the back of the tomato at the same time amounts to representing the whole object. I think not; representation of both the visible and the occluded parts is not sufficient for the representation of the whole object. This is because representation of the whole object is about objecthood, while representation of the front and the back is about the shape and the color of an object. You might represent the conjunction of detached visible and occluded parts of an object without representing them as parts of an object as in the classic Quinean case. Thus, the representation of both the occluded parts and non-occluded parts are not sufficient for the representation of the whole object. However, the representation of both the visible and occluded parts seems necessary. In his review of *Action in Perception*, M. G. F. Martin (2008) argues that we visually experience the whole object by virtue of experiencing the front, denying any kind of visual experience of the back. I think Martin is right in that we visually experience the whole object in virtue of experiencing the front, but I think he is wrong in his denial of the experience of the back. It seems to me we experience the back in virtue of experiencing the whole object. For example, representing a mug as a mug enables you to have a general idea about how the back of the mug would be like; it would have a mug-like cylindrical shape rather than being completely flat. If the mug were not represented is as a whole, there would be infinitely many possibilities what the back might be like—it might even lack a back! In other words, the representation as of a whole object works as a "scaffold" for what the back of the object is like. I will motivate this tripartite account more in the end of this section.

Keeping these three kinds of representation and the relationship between them in mind, let's review other instances where the notion of perceptual presence is invoked.

> Presence in absence
> When you hold a bottle in your hands with your eyes closed, you might feel the presence of the entire bottle.

Here, you feel the sense of presence of the whole bottle even though you are directly only sensing the part of the bottle you are touching. This case is different to the previous cases in that no occlusion has taken place, however, it is similar in that it is also a case of experiencing world of objects despite informational limitations. In occlusion, sensory information from the certain parts of the object does not arrive at one's sensory organs because of the occlusion. Similarly, in presence in absence, sensory information from the untouched parts of the object is not registered because parts have to be in contact with a hand in order to register haptic information. In spite of the absence of information from the untouched parts, your perception is as of the bottle, rather than a mere aggregation of the parts[16]. Consideration of this example reveals that representing the whole of an object without receiving sensory stimulation from all its parts, and representing the parts without receiving sensory stimulation from them is a key characteristic of perceptual presence. Hence, I shall call this category of perceptual presence *presence in absence*, by which I mean representation in the absence of sensory stimulation.

Another kind of perceptual presence is modal completion,[17] such as Kanizsa's triangle. Kanizsa's triangle (Figure 6) is usually perceived as an illusory triangle partially occluding three disks, rather than as three pacmen facing a central point. That is, we can interpret the figure as a case of occlusion.

> Modal completion
>
> [W]e naturally perceive this figure as the depiction of a triangle partially occluding three disks. We don't merely think the presence of the occluded bits. (Noë, 2004, p. 61)

This type of case is a bit different to straightforward occlusion because it involves an illusory contour. However, it can also be considered in the same way as occlusion cases. This case still involves representation of the whole object (representation as of the disks) and representation of occluded parts (occluded parts of disks); we represent the disks as complete circles partially obscured by the triangle, and not as pacman-shapes. You might doubt that we actually *perceive* the hidden contour, and think instead that the

[16] Surely, for this to occur, you need to have previously experienced holding a bottle; you need to have prior expectations for the bottle. The trickier point is whether sensory modality of acquaintance in advance with a bottle matters. Someone who has had prior visual and tactile acquaintance with an object might have different expectations from someone who has only had tactile acquaintance. This is a version of Molyneux's problem, which I will not pursue here.

[17] Noë introduces Kanizsa's triangle as an instance of *amodal* perception, but it is usually treated as modal. The difference between modal and amodal completion is relevant to the topic of this chapter. In modal completion, an illusory contour is perceived as if it has the same *mode* as the actual contour. An example is the contour of the triangle in Kanizsa's triangle. But in amodal completion, a contour is not perceived, at least, not in the same sense as a visible contour. A typical case of amodal completion is occlusion. You don't see the cat's hidden tail in the same sense as you see the visible side of the cat. However, as discussed, you may interpret the Kanizsa's triangle as involving occlusion if you talk about the disks, rather than the triangle.

contour is merely *judged* to be there. Regarding this, Noë (2004) makes an interesting argument from phenomenology. He argues that the contour is perceptual because it feels different from when you merely consciously think of a triangle and three occluded disks from phenomenological perspective. But this argument from phenomenological difference is not conclusive; it might be explained differently from a position that explains the hidden contour of disks by positing some cognitive state.

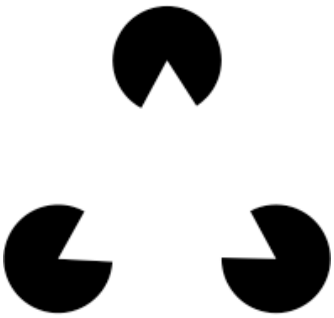


Figure 6 Kanizsa's Triangle

This cognitive view might try to account for the phenomenological difference between these two experiences. When you are merely feeling the presence of the contour, only two representations are responsible for your experience—representation of visible sides of disks (perceptual) and that of hidden contour of disks (cognitive), but when you consciously judge there are a triangle and three occluded disks, you get a new representation on top of those representations at play; there are three representations now: the perceptual representation of the visible side, the cognitive representation for the hidden contour, and the conscious representation of judgment. This difference might account for the phenomenological difference.

Evaluation of this argument requires principled treatment of the familiar cognitive/perceptual distinction. This will be given by the predictive processing account, and will therefore be discussed in Section 4.

The next case involves attention.

> Perception without attention
> [For] example, I may look at you, attending only to you. But I also have a sense of the presence of the wall behind you in the background, of its color, of its distance to you. […] we must explain how it is we can enjoy perceptual

experience of unattended features of a scene. (Noë, 2004, p. 59)

In this case, Noë talks about perception without attention. When you consciously attend to an object, everything else seems to fade from consciousness. However, Noë's point is that we still perceive *something*. Unattended areas are not totally blank; unattended objects do not seem to be nonexistent. We might not see the wall in great detail but we still have some ideas about the color of the wall and the distance of the wall from us. Although this case can be treated similarly to the previous instances, it has an important difference. In the foregoing cases, there is no sensory stimulation from the relevant parts. However, in perception without attention, there is sensory stimulation to the retina, even if that stimulation is minimal. So now the problem is cast in more comparative terms; there is more to experience than simply information to the retina. Presumably because of this, people like Prinz discussed above, might be less reluctant to admit this is perceptual. There is nothing to prevent objects from being visible, it is just there is not as much sensory stimulation as in fovea. Still, what makes this case a case of perceptual presence is that we experience objects even though there is not sufficient information to represent fine-grained details such as color or shape.

Presence, as Noë talks about it, is not restricted to perceptual cases; there is actually *nonperceptual* presence. Those cases are still cases of presence because you feel the *existence* of what is not directly perceived just as you feel the presence of the back of a mug. But in this case, your feeling of presence is not perceptual.

> Nonperceptual presence
> You also have a sense of the room next door, for example. But your sense of its presence is not a sense of its perceptual presence. It doesn't seem to you now, as if you see the space on the other side of the wall. (Noë, 2004, p. 64)

Noë gives his own account of the difference between perceptual and nonperceptual presence (Noë, 2004, p. 64). To explain the difference, he invokes two kinds of sensory relation between a perceiver and the world. The first is movement-dependence. If the slightest movements of the body modulate sensory stimulation from an object in the world, then the relation between the perceiver and the object is movement-dependent. It seems almost trivial that perceptual relations are movement-dependent. For example, when you move your face, your visual experience of a refrigerator in front of you varies. The sensory stimulation from the refrigerator changes, depending on the angle and distance between the eyes and the refrigerator. However, when we think about a chair in the next room, almost all possible actions of a perceiver do not give rise to change in sensory stimulation, simply because there is no sensory stimulation arriving from the object. The exception is you're actually going into that room. However, apart from the small subset of actions that does give rise to the reception of sensory stimulation from the object, movement does not give rise to sensory change. Therefore, Noë thinks non-perceptual relations are *less* movement-dependent.

The second kind of relation between a perceiver and the world, according to Noë, is object-dependence. If movements of an object produce sensory change, the relation between the perceiver and the object is object-dependent. Again, it seems clear that most movements of an experienced object produce sensory change. The movement of a car, perceived on a street, would produce change in your retina and visual experience. At the same time, most movements of unexperienced objects do not give rise to changes in sensory stimulation, apart from the case of an object starting to be experienced (such as when a person appears from the edge of your visual field). Thus the relationship between the perceiver and experienced objects is both object-dependent and movement-dependent.

Given this distinction, Noë argues that non-perceptual presence is not object-dependent and is less movement dependent than perceptual presence. More precisely, the relationship between the perceiver and the objects that are the target of felt presence is not object-dependent nor movement dependent. In the case of the felt presence of the room next door, movement of an object in the room would not, except in extreme cases, register in any of your sensory organs. Furthermore, even though you can move to the room and see what's there, again, most of your possible actions would not cause you to experience any sensory stimulation from the room next door.

I think Noë's proposal is helpful here to think about what perceptual presence consists in. The two kinds of relationship discussed here reveal that perception of an object is an essential element of perceptual presence. This is because the relationships are ones that obtain between the perceiver and objects and are defined in terms of how changes in those two relationships give rise to change in sensory stimulation. Noë's proposal is that objects are not seen if any movement on either the perceiver's part or the object's part does not produce sensory change. Conversely, when change is produced, we perceive the whole object not only aspects of the object. This is compatible with the definition of perception given in the preceding chapter (p. 35).

> *Perception in predictive processing*
> The content of perception is the content of models at all levels that are used to explain away sensory stimulation

According to this characterization, a representation of an object is included in perception only when it is used to explain away sensory stimulation. If not, it is not perceptual. This is virtually identical to Noë's claim above about the distinction between perceptual and nonperceptual presence in that both accounts posit that an object needs to influence sensory organs to be perceived. Paradigmatic cases of perceptual presence can also be partly explained in terms of perception of objects. In those cases, at least some parts of an object are visible to the subject (i.e. sensory simulation is arriving at the retina), so a change in that object can easily be reflected in sensory stimulation (e.g. rotating a mug in front of you). Moreover, when you see a mug, your representation of the front of the mug gives rise to expectation for the representation at the object level of the mug, and this in turn gives rise to an expectation regarding what the back of the object looks like. Put more generally, the representation of the visible parts determines what object the perceiver is seeing, and the representation of the whole object finally leads to the formation of a representation of the back. Furthermore, the proposal implies that the perceptuality of presence is a matter of degree because both object-dependency and movement-dependency come in degrees.

The case of the cat whose body is partly occluded by a picket fence is a typical case of *perceptual* presence and it is true it is more object-dependent and movement-dependent than the presence of the room next door. If the cat moves forward, your experience of the cat will change, and if you move, that would also change your experience, most of the time. However, there are some ways in which the cat can move that won't impact your sensory organs. For example, suppose the cat's tail is perfectly occluded by the fence. In this case, any movement of its tail would not cause any change in the pattern of stimulation you get from the cat, and yet you'd be surprised, absent prior knowledge, if the cat was tail-less. So, in a nutshell, the more object and movement dependent presence becomes, the more presence becomes perceptual.[18]

One might wonder what kind of representation is involved in the case of nonperceptual presence if it is not perceptual. Noë doesn't give any positive representational characterization for nonperceptual presence, probably because he is an anti-representationalist. But it seems natural to think this is a case of cognitive representation, as it is not related to any immediate sensory stimulation from a representationalist perspective. This consequence resonates with my own conclusion developed in sections 3 and 4.

By examining a variety of cases of perceptual presence, we can arrive at its core. The problem of perceptual presence is, in a nutshell, the question of how we experience the world of objects from our limited contact with those objects. We represent objecthood through representing the parts of an object that can stimulate our sensory organs, and this representation in turn, allows us to represent the parts of an object that cannot stimulate our sensory organs. In any case of perceptual presence, at least some parts of a given object can stimulate our sensory organs and that sensory stimulation sets constraints on what the object is and what the parts of the object that cannot stimulate our sensory organ are like.

Before moving on to the next section, it will be useful to revisit the three kinds of representation. The first was the representation of the front of the object, which can now be characterized more generally: the representation of the parts of the object that stimulate our sensory organs. The contents of the representation are perspectival: perceived aspects change when the spatial relationship between the perceiver and the object changes.

The second kind is the representation as of the whole object. The experience is non-perspectival in nature; no matter which angle you see an apple from, you see "the apple." Compared to experience of the front of the object, the phenomenology of experiencing the whole object is not easy to grasp—it is something stable across different perspectives. Because of this insensitivity, one might disagree that this

---

[18] We can tell a similar story also on examples of other sensory modalities. Consider a tactile example. Suppose you are blindfolded in a room but you remember where the entrance door was—you feel the presence of the door but it is clearly not perceptual. And in the same situation, you also feel the presence of the whole floor due to the feeling of your feet, but it might be less perceptual than the bottle case because you only touch a small region of the floor.

representation is perceptual in nature and claim instead that all that exists is the judgment that this is an apple. This is the problem of representation—what kind of representation is the representation of the whole object?

The last kind is the representation of the back of the object, now characterized as the representation of parts that do not stimulate one's sensory organs. The experiential nature of the representation is also perspectival. As you move around a still object, the parts once visible become hidden, and vice versa. Again its phenomenology is again far subtler than that of the front representation, but its subtlety seems different to that of the representation of the whole object. What had been in question about the representation of the whole object was its perceptual nature because of insensitivity to perspectives, whereas what is at issue in the representation of the back is rather its existence itself. Even though the representation of the back has perspectival content in that it concerns shape of an object seen from an angle, its phenomenology is faint. Thus, Skepticism about the existence of the representation of the back is not unreasonable. For example, Martin (2008) claims we perceive the whole object in virtue of experiencing the front but denies that we have any experience of the back. Those who think we experience the back can counter that the phenomenology of the experience of the back is faint but nevertheless exists. Appealing to empirical enables progress in this otherwise infertile debate. In the literature on amodal completion, neural activity that corresponds to the occluded contour is found (Shimojo & Nakayama, 1990; Watanabe, 1995). This suggests we have at least something akin to visual representation of the illusory contour. Thus, there is at least some reason to think all of the three kinds of representation are actually involved in perceptual experience. However, the problem of representation remains. A suitable account has to explain these three kinds of representations and their (non-)perspectivalness, their vividness, and their subtlety (See Table 3 below).

Table 3 The three kinds of representation involved in perceptual presence

Representation	**Perspectivalness**	**Clarity**
Representation of the front	Perspectival	Yes
Representation of the back	Perspectival	No
Representation of the whole object	Non-perspectival	?

3.3. *Nanay on amodal perception*

In the previous section, I reviewed instances of perceptual presence. In this section, I examine Bence Nanay's (2010) paper on the problem of representation. He specifically discusses amodal perception but I will generalize his argument to perceptual presence in general using the analysis given before. Nanay addresses how we represent the hidden contour in the case of amodal perception and argues this is through imagery. He considers four options: a perceptual account, a belief account, an access account, and an imagery account; and he concludes, after eliminating the other three options, that our representation is best explained by the imagery account.

Nanay first discusses the perception account, according to which we actually see the occluded parts. But he rejects it because there is no relevant sensory stimulation and there is no activation of the cells in the retina corresponding to the perceived sides of the triangle in Figure 6. However, I think this rejection is too hasty. Many philosophers and psychologists including Noë and Gibson take the perception account seriously despite the fact that it is surely obvious to them that there is no sensory stimulation from the occluded parts. Perhaps, we need other reasons to think it is perceptual apart from phenomenology, but rejecting this option simply because of the lack of sensory stimulation seems too quick. I will investigate positive reasons for this option in Section 5.

The next position Nanay criticizes is the belief account, according to which the representation is somewhat belief-like. Apart from the phenomenological objection that amodal completion does not feel like a typical belief, he gives two original objections. The first is that the representations are insensitive to what you believe. The second is that empirical findings suggest that the representation is instead sensory. As I mentioned before, activation patterns that correlate to the percept are found as early as the primary visual cortex.[19] However, Nanay's rejection again seems too hasty to me. Even though it is true that no matter what you believe about the objects in Figure 7 below, their appearance doesn't change, there is a point in regarding the representation as belief-like because there seem to be elements of inference. The shapes represented in

[19] One way to explain the findings from the belief account is to regard the activation as top-down influence from belief. Nanay argues this is unlikely, evidence for this position can be found in Gilbert and Sigman (2007).

amodal perception are actually ambiguous between many hypotheses. For example, in Figure 7, a circle is usually perceived as being behind a square, but the figure occluded by square could instead be a "pacman" as in Kanizsa's triangle, or it could in fact be any shape that could be occluded by the square. The brain somehow "infers" the most likely cause of the percept and thereby we perceive a circle. That is, we have some background beliefs about what things are like in the world.

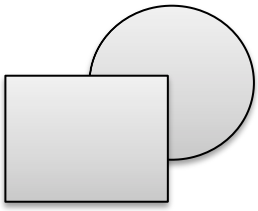


Figure 7 An example of modal completion

The third account rejected by Nanay is the access account, which is Noë's view.[20] According to the access account, we *do not* actually represent the occluded parts. Rather, we only have access to the occluded parts and that explains the sense of presence. Nanay points out that this view has difficulty explaining the difference between perceptual and non-perceptual presence. Even in the case of the presence of the next room, which is a nonperceptual case, the subject has access to the room next door. However, as we have seen, Noë gives an account for the distinction between perceptual and nonperceptual presence referring to object-dependence and movement-dependence. There also is another challenge for access account raised by Clark (2012). The access account, and indeed any theory that does not posit a representation in the head, struggles to tive an explanation of the similarities between perception, imagery, and dreams. Perception, imagery, and dreams are phenomenologically similar. For one thing, all of them are intentional states. Perceiving a sports car, visually imagining a sports car, and dreaming a sports car are all about a sports car. Moreover, they all accompany sensory phenomenology. This similarity can be easily accounted for by positing that all have the same form of representation with contents. However, this route is closed to an account,

[20] As I mentioned, Noë is an anti-representationalist and conceives of perception in terms of *access*. Therefore, Noë's own view is an access and perception account.

such as the access account, that denies the existence of representation in the brain.

The last account Nanay considers is the imagery account. This is the view he defends. What he means here is *visual imagery*, a quasi-perceptual state occurring without appropriate sensations.[21] Classifying amodal perception as a kind of imagery helps to accommodate its similarity with perception both in phenomenological and neural terms. As discussed above, you can find brain activation as early as primary visual cortex and it is known that the same neural areas are activated both in perception and visual imagery. Nanay anticipates objections that point to dissimilarities between standard cases of imagery and amodal perception. Imagery usually cannot be located in actual egocentric space. For example, when you imagine you play in a professional baseball team, the imagery cannot be located in your current egocentric space. However, according to Nanay, imagery can be localized. You can imagine there is a succulent steak on the table in front of you, for example. In this case your imagery of a steak can be located at a specific location on the table. Another possible objection Nanay anticipates is the effortlessness of amodal perception. Typical imagery is effortful, as when you visualize a steak in front of you. If you imagine a steak without any effort, it might rather be considered as perception or other mental states. On the other hand, amodal completion is not effortful; you don't try to see the hidden contour in amodal perception, the hidden contour just appears to you spontaneously. Nanay thinks this objection is based on the mistaken idea that imagery is always effortful and attentive.[22] Nanay appeals to inattentional blindness here. Inattentional blindness is the thesis that we are not aware of things we are not attending to. Nanay argues, on this basis, that we don't notice the existence of non-attentive forms of imagery because of inattention, and perceptual presence is precisely that kind of imagery.

However, if Nanay's explanation is correct, we constantly engage in imagery when perceiving. Nanay tries to stretch our ordinary conception of imagery, but he goes too far. First of all, it is not clear if the appeal to inattention justifies his claim. His claim is that we are not aware of the nature of events (i.e. imagery) because of inattention. In

[21] He does not take a stance on the famous debate about the representational format of mental imagery (whether it is pictorial or not). Pending its representational nature, he refers to a representational state that gives rise to quasi-perceptual phenomenology.

[22] However, many philosophers think that effortfulness is hallmark of imagery. See Ichikawa (2009) and Wittgenstein (1967)

inattentional blindness, we are not aware of external objects in the world if attention is not paid to them. And if we are not aware of objects, it is reasonable to think we do not know the nature of the representation of those objects. But in the case of amodal completion, we are aware of the contour and we are also aware of ourselves completing the contour somehow. So the parallel between inattentional blindness and amodal completion is imperfect. Moreover, it is mysterious to suppose we do not know anything about the nature of completion because of inattention, if you consider the fact that we are aware of the contour and of our own experience of that contour.

I think it is natural to suppose we cannot classify the nature of completion correctly mainly because there is no folk-psychological category that perfectly fits completion. For one thing, when completion is represented, it has some top-down elements; it is reminiscent of more cognitive states. For another, the associated phenomenology is so subtle that ordinary people might not have needed terms to talk about the phenomenon. Secondly, if you can stretch a folk-psychological concept so as to include cases of completion, you can stretch any folk psychological concept so that amodal completion can fit. For example, you can stretch the notion of belief so as to incorporate opposing beliefs in some occasions. Moreover, classifying the representation of the hidden contour as imagery misses the interplay between the representation of visible parts, the representation of the hidden contour and the representation as of the whole. I will address the interplay in section 4.

Lastly, Nanay's account is unsatisfactory as a full account of perceptual presence because he only addresses the representation of occluded parts in amodal presence. There is still the need for an explanation of representation as of the whole object.[23]

I have argued that Nanay's argument fails because the representation of the hidden contour doesn't sit neatly with any of the folk-psychological conceptions he discusses. In the next two sections, I will try to sketch a mechanism that underlies perceptual presence under predictive processing. As I argued in chapter 1, predictive processing gives a holistic and revisionistic account of the mind. By giving an account of perceptual presence in terms of predictive processing, we can see why the relevant representation did not fit any folk psychological categories.

---

[23] If an imagery account can explain the representation as of the whole object, it is by way of treating imagery as prediction from the representation as of the whole object. But this is tantamount to the account developed in the next section.

3.4. *Predictive processing and perceptual presence*

In this section, I will explain the three kinds of representation from the perspective of predictive processing, as introduced in Chapter 1. The tools used to explain perceptual presence will be *Prediction Error Minimization*, *Hierarchy*, and *Precision Expectation* (these were explained on pp. 26-30).

According to *Hierarchy*, the system implemented by the brain has a hierarchical structure. Each level predicts bottom-up signals coming from the level below, and what cannot be predicted at that level is the prediction error, which is sent to the level above. At the bottom of the hierarchy is the sensory signal to be explained away by top-down prediction. We can situate the three kinds of representation in this hierarchical prediction error minimization system, and use the characters of hierarchy to account for the representations. I will argue that hierarchy is regimented according to timescale of causal regularities and because of that, lower levels concern local, fine-grained, and more perspectival aspects of the world whereas higher levels concern global, coarse-grained, and less perspectival aspects.

Given this characterization, how can the three representations be located? Firstly, the representation of visible parts is located at a lower level than the representation of the whole object (Clark, 2012).[24] The visible parts of an object are certainly more local than the whole object. Moreover, objecthood is subject to a longer timescale of causal regularity. The mugness of a mug cannot be revealed instantaneously—visible parts can be confirmed just by having one percept (hence visible), but confirming that it is a mug requires viewing the object from many angles. Some viewpoints might be more important than others. For example, viewpoints that include the handle would be more critical than other angles, but they are still not conclusive by themselves. You have to check that the object has a bottom, for example. Moreover, objecthood is less sensitive to details. For example, a mug can take various shapes and colors. Lastly, the representation of occluded parts of an object seems to be located at the same level as that of the visible parts. Both visible and occluded parts belong to the same genus: representation of color and shape.

The interplay between lower-level representations and higher-level representation can be explained in predictive processing terms. Prediction errors at lower-level are sent to the higher-levels and the higher-level representations can explain away the errors by disambiguating the lower-level representations (Yuille & Kersten, 2006). Sensory stimulation is often ambiguous between multiple hypotheses (e.g. whether this contour is continuous with that contour, whether this surface color is same as that surface color). To address this, higher-level models can provide contextual bias towards certain hypotheses at lower levels by implementing top-down predictions. In the context of perceptual presence, a higher-level representation (e.g. the representation of a whole cat) makes a prediction about the lower-level representation (e.g. the representation of the shape and the color of the cat). The representation of visible parts (e.g. the representation of the parts of the cat) emerges as the winning hypothesis as its shape best explains the sensory stimulation. However, the representation of unseen parts is also predicted by top-down prediction from the higher-level model (e.g. the shape of the tail of the cat occluded by a fence). This representation of the unseen parts is not assigned higher posterior probability, because it is inconsistent with incoming sensory stimulation (i.e. relevant sensory stimulation is not obtained because of occlusion). The subtle phenomenology of unseen parts may be thus explained by this lower posterior probability.[25] The representation of the unseen parts also lacks details. Recall that higher-level representations carry information in less detail. Thus the representation of unseen parts as a prediction from a higher-level representation also lacks detail. This makes intuitive sense: when you see a mug and you think you know what the back of the mug is like, you only have expectations about its shape, consistent with what is visible and what you know of the shape of ordinary mugs. You have no idea about the specific shape, nor any possible chips or distortions it might have.

To summarize, the representation of the front is a representation at a lower level, whose contents are perspectival and are richer in detail. The representation as of the whole object is a representation at a higher level, whose contents are non-perspectival and less rich in detail. Finally, the representation of the back is again at a lower level,

[25] Madary (2015) argues that the determinacy of perception corresponds to probabilistic coding in predictive processing. Is this relevant here? I think it corresponds to different aspects of probabilistic coding. Whereas determinacy directly corresponds to the precision of the model, subjective clarity of the model rather corresponds to probability of the model.

but since it is a prediction from the representation at the higher level and the prediction does not involve sensory stimulation, the representation is not assigned a high probability. This explains why its phenomenology is subtle. This is a case where the probabilistic nature of predictive processing plays an important role.

In this section, I located the three kinds of representation using the toolkit of predictive processing. This illuminates the interplay between them. The hierarchy of the brain explains the difference between the representation of the front of something, and its representation as of an object. And prediction from higher-level processing explains the representation of the back, and why this representation lacks details.

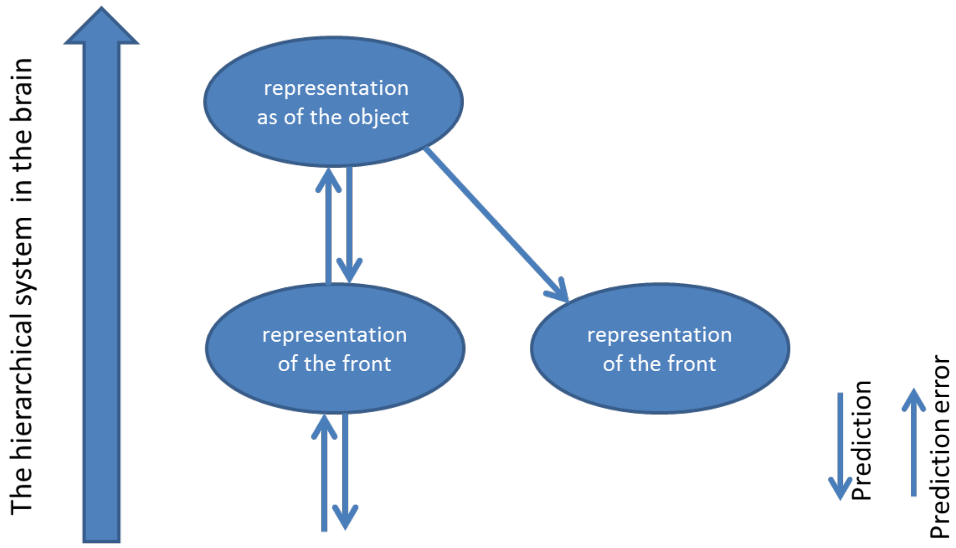


Figure 8 Three representations located in the hierarchical brain of predictive processing

3.5. *Solving the four problems*

As I argued above, perceptual presence actually consists of two problems: one is the problem of representation and the other is the problem of subjective veridicality. I will answer these problems along with the three other problems (inverse, description, and reason) I set out in Chapter 1. I will start with the problem of representation, then proceed to three problems general problems, and then finish with the problem of subjective veridicality.

Answering the problem of representation

In this section, I give a solution to the problem of representation based on the predictive processing account given earlier. First, let us recall what perception is under predictive processing:

> Perception in predictive processing
> The content of perception is the contents of models at all levels that are used to explain away sensory stimulation

I also characterized non-perceptual mental states as those that are not in the business of explaining away current sensory stimulation. Under this framework, how can we classify perceptual presence?

Let us start from seeing how this treatment of perception differs from Nanay's. Nanay argued that (successful) perception is caused by sensory stimulation. This point is also granted by the predictive processing account as evidenced by the definition, but the resultant content of perception might not be the exact cause of sensory stimulation. As I argued in chapter 2, one of the fundamental ideas of predictive processing is that sensory stimulation underdetermines its cause, and thus the brain has to infer based on priors. Moreover, depending on context, the precision of prediction error might be deemed low, and thus perceptual content might rely more on priors (e.g. when the sensory stimulation is expected to contain much noise). Thus, under the predictive processing account, even though the models are about the sensory causes, the content of the models can go beyond the information obtained through sensory stimulation. If this point is granted, we can argue that *all* three representations in perceptual presence are perceptual representations. All of the representations are the result of explaining away sensory stimulation. Even the representation of unseen parts is the result of prediction error minimization. Due to the lack of relevant sensory stimulation, it is predicted from the higher-level representations.

In Chapter 1, I argued that elements of the higher-level representation, such as that of the whole object, have similar characteristics to concepts but that this does not exclude such representations from being a form of perception. As I argued, the higher-level representation of the whole object plays an indispensable role in explaining away prediction errors at the lower level, and so it is part of perception. As I argued in Chapter 2, conscious thoughts may also be made up of higher-level representations but conscious thoughts are not usually in the business of explaining sensory stimulation.

Contradicting higher-level representations are not employed at the same time in either perception or thoughts. The apparent contradicting case is the case where one representation is used in perception and the other is not. In amodal completion, you might think, “there is no object that the contour belongs to” while you are representing the whole object completing “hidden” contour. In this case, only the latter representation is used to predict sensory stimulation.

To summarize, I argue the representations in perceptual presence are actually perceptual. Representations at different levels have different characteristics—some deal with longer timescales and therefore exhibit similar characteristics to concepts, some are the opposite. However, they can all be forms of perception if they are all called upon to make sense of sensory stimulation.

Nanay’s motivation for arguing that amodal completion is represented as imagery comes from Neo-Kantian view on perception such as Strawson’s (1974). Strawson argued that imagery is an essential element of perception. Following this line of reasoning, Nanay argues that an imagistic, endogenous, or constructive component *different from normal perception* is involved in representing amodal completion or the perception of unseen parts in general. However, under the predictive processing framework, *every aspect of perception is imagistic or constructive* in the sense that perceptual content is generated by representation stemming from top-down prediction. Thus, there is no difference between the representation of visible parts and that of unseen parts. Because of this, the predictive processing account emerges as the simplest, and also fully Kantian view among all.

Answering the inverse problem

The inverse problem in perceptual presence can be stated as follows: how can we perceive the whole of an object even though the visible parts are compatible with many candidates? For example, the object you are seeing right now could be a half-mug, completely lacking a back. But we see the object as a complete mug and would be surprised to find it lacking aback. How is it possible? In a Bayesian framework, this is simply the result of learning from “data.” That particular pattern of sensory stimulation is typically caused by a complete mug in this world rather than by a half-mug even though both possibilities explain the sensory stimulation. As I have explained in Chapter 2, this is the “best guess” from current sensory samples, even though it can sometimes go awry.

Answering the description problem

There are many disagreements regarding the description of perceptual presence. The most significant of these regards the representations involved in perceptual presence. We saw Prinz and Martin have even denied the existence of a representation of the unseen parts of an object. Predictive processing account explains why its phenomenology is so faint that even its existence is doubted.

Answering the reason problem

Why do we experience perceptual presence at all? In the inverse problem, it was questioned how the experience of the whole object is possible. But if you think about the case of the Simpleton in Chapter 1, the most important and primary thing for a biological system is to distinguish surrounding objects. Thus, the problem can be looked from the other side; why do we experience perspectival aspects at all? The answer is found in the inverse problem: sensory stimulation from an object is ambiguous at many levels. We see that in a hierarchical setting, the higher-level models can help dissolve ambiguity at lower-levels. Representation at multiple levels is the best solution to solve the inverse problem and this mandates perceptual presence and perspectival aspects.

Answering the problem of subjective veridicality

Now it's time to turn to the final problem: the problem of subjective veridicality. In his recent paper, Anil Seth (2014) addresses this problem. His theory is a fusion of Alva Noë's enactivism and predictive processing. According to enactivism, the mastery of sensorimotor contingencies explains perceptual presence. Sensorimotor contingencies are patterns of change of sensory stimuli when a subject acts in a certain way. When a subject "knows" how the sensory stimuli changes given an action, she enjoys a certain perceptual presence. As I have argued above, Noë's own formulation of the problem of perceptual presence is ambiguous between the problem of representation and the problem of subjective veridicality. Seth specifically focuses on subjective veridicality and takes Noë as providing solution for that problem. Seth doesn't take issue with this hypothesis itself. Instead, he tries to specify the neural mechanism that underlies these sensorimotor contingencies under predictive processing. By doing so, he claims that perceptual *absence* of hallucinatory perception in synesthesia (a typical subject of synesthesia does not feel like his or her experience is real) can be accounted for. Synesthesia is a condition in which a stimulus (inducer) causes another extra experience (concurrent) in a subject. For example, hearing a certain tone can cause a color experience in addition to normal sound experience. It is a common feature of synesthesia that even though their experience of concurrent (e.g. induced color experience) is vivid, it does not feel real (Seth, 2014).

In the standard predictive processing story, the hierarchical system is employed to explain current sensory inputs. Seth accommodates the mastery of sensorimotor contingencies in enactivism by allowing representation in the hierarchical system to encode *counterfactual* predictions between action and sensory inputs. This is a direct incarnation of sensory motor contingencies right into predictive processing, so he dubs his theory the theory of predictive processing of sensorimotor contingencies (PPSMC).

Equipped with this counterfactual element, Seth explains the difference in subjective veridicality between experiencing inducers (and many other ordinary experiences) and experiencing concurrents. That is, representations of concurrents at intermediate levels in the visual hierarchy are counterfactually poorer than those of inducers at intermediate levels. Inducers exist out there in the world, and so there are many sensorimotor contingencies to be learned, but concurrents are actually not in the world so there are few sensorimotor contingencies to be learned. This difference yields the difference in perceptual presence.

Hohwy (2014) argues against Seth's account and claims that nonperspectival representations of an object are sufficient for perceptual presence.[26] This can work as an answer to the question about the kinds of representation involved in perceptual presence, as I have argued that perceptual presence involves three kinds of representation. However, invocation of nonperspectival representation seems an unsatisfactory answer to the question of why it feels real. As mentioned before, patients suffering from Charles Bonnet syndrome are subject to visual hallucinations are otherwise cognitively healthy.[27] The nature of these visual hallucinations was first described by Charles Bonnet, a famous Swiss philosopher and naturalist, whose

[26] Clark (2012) also develops a similar line of argument.
[27] Subjects also typically have a problem in visual acuity (Lerario et al. 2013)

grandfather who suffered from the syndrome:

> For example, I therefore limit myself to saying that I know a respectable man, full of health, of innocence, judgement and memory, who, completely watchful and independently from all outside influences, perceives from time to time, in front of him, Figures of Men, Women, Birds, Vehicles, of buildings. He saw these figures make different movements… He saw the Tapestry of his apartments suddenly change, in tapestries of another flavour, richer… The person I am talking about was subjected more than once and in old age to cataract surgery on both eyes… (Bonnet, 1760, as cited in Lerario, Ciammola, Poletti, Girotti, & Silani, 2013)

The contents of these hallucinations can be simple, but can also be as complex as lifelike scenes. Importantly, people with Charles Bonnet syndrome usually "have insights into the unreal nature of their perceptions" (Lerario et al., 2013, p. 1181). It is difficult to understand, from "have insights", whether this refers to a more spontaneous felt "unreality" or to a more cognitively understood "unreality." However, in at least one case study, there seemed to have been felt unreality because the recognition had been acquired instantly at the onset of hallucination, as Lerario et al. (2013) write, "He found the images quite upsetting, but retained insight throughout". Thus such representations need not correspond to a felt reality.

It is reasonable to think that if a hallucinating patient is subject to a complex life-like scene, objects in that scene will be represented as objects, and thus nonperspectival representations are activated despite them feeling unreal. Thus, something more is required to solve the problem of subjective veridicality.

Hohwy (2014) also questions how the counterfactual elements can give rise to ocurrent phenomenology of presence. To see if Hohwy's criticism is fair, we need to revisit Noë own argument for sensorimotor contingency because Seth only employs Noë's hypothesis about perceptual presence and builds it into his own account. Noë's (2004) argument has two steps. First, he claims that careful phenomenological analysis reveals that no representation is involved when we feel the presence a partly occluded object, such as the whole cat behind a picket fence. Second, Noë argues that access to the cat is a matter of having appropriate sensorimotor contingencies. Here, sensorimotor contingencies are introduced to explain the ocurrent access that a perceiver has to the cat. It is certainly true that sensorimotor contingencies are about counterfactuals. However, sensorimotor contingencies, and the access enabled by them are actual (as opposed to merely potential) skills a perceiver possesses. Therefore, Hohwy's criticism that counterfactual knowledge cannot underlie actual concurrent phenomenology is not necessarily valid.

However, Noë's approach comes at a cost; he has to admit that no representation is involved in perceptual presence. This doesn't seem right. As I argued in Section 3.2, explaining perceptual presence in terms of three kinds of representation is a better account than the access account which doesn't posit any representation. Moreover, the access account doesn't seem a good account also for the problem of subjective veridicality. There are forms of experience in which a subject enjoys presence but does not have any access to objects: dreams or perceptual hallucinations. In those kinds of experiences, a subject cannot have access because the objects don't exist!

I think a more plausible way to incorporate sensorimotor contingencies or some counterfactual element in predictive processing is to consider them as constitutive elements in high-level object representation rather than incorporating them at the intermediate level representation as Seth proposes. I previously explained how high-level representations make predictions about intermediate level representations, such as what the unseen back of an object is like. This regards the *current* perspectival shape, but if this is possible, it seems plausible to think a high-level representation could also make conditional predictions about how an object would look given a certain action. After all, the shape of the back is what you will see when you turn the object around! To put it in plainer terms, if you know what kind of the shape a mug has, you know how it would look when you view it from a certain perspective. It would look circular when viewed from the top, when viewed from the side that does not have a handle, it would look as if there is no handle, and so on. The high-level representation might even enable the brain to predict the ideal perspective a viewer should take to have the best chance of correctly identifying a mug. The high-level model should carry information about core characteristics of a mug. For example, there is need to distinguish between a mug and a cup.[28] To distinguish between them you might want to get closer to the object, for

[28] The distinction is not clear-cut, yet there are paradigmatic mugs and cups, with

example. As Seth himself admits, this aspect of high-level representation is already incorporated in *Active Inference*. According to Active Inference, the subject acts based on a high-level model to search for congruent sensory stimulation that implies there is information about how to act and what to expect after action is included in the model.

This way of incorporating counterfactual elements enriches my account given in the Section 4. Some might wonder if a higher-level representation of the whole object is sufficient to explain the phenomenology of unseen parts. I don't think that is the case. My reason for this disagreement is close to Hohwy's criticism. The unseen parts are not experienced as counterfactual elements. Rather they are rather given as something that exists "right there" even though they are unseen. Moreover, this counterfactual explanation does not explain why the phenomenology of unseen parts is subtle. Therefore, my account of representation of unseen parts stays truer to phenomenology.

If counterfactual richness does not explain subjective veridicality, what can do that? To see that, I think we should go back to the difference between what feels real (inducers) and what do not (concurrents). Inducers and other ordinary objects interact each other in law-like manners. For example, suppose a mug is partially occluded by another mug. As you move around these mugs, their relationship changes and the visible parts of them change relative to the position of the viewer. On the other hand, concurrents do not interact with other objects. Their appearance is remains constant no matter how the other objects are arranged or how a viewer moves. The brain learns this difference in interaction and learns that the concurrents are not real. Seth focuses on this same point. He says, "The hidden causes giving rise to concurrent related sensory signals do not embed a rich and deep statistical structure for the brain to learn" (Seth, 2014, p. 108). Put in simpler terms, since concurrents do not actually exist in the world they cannot interact with other objects in the world. Thus sensory signals do not show complex action-dependent patterns, and so there are no sensorimotor contingencies to be learned.

The main difference between Seth's account and mine is that I don't cash out the difference between inducers and concurrents in terms of counterfactuals at the intermediate level representation. I think the absence or presence the perceived object in the real world has to be learned by the performance of actual actions, at least at the time

---

differences between them.

of first exposure to a given type of experience. Only through these actions does a subject learn the difference between something real and something unreal. Also, the representation related to subjective veridicality sits at a higher level because the representation involves the relationship between two or more objects and the relationship between objects and the subject. Thus this would be processed at a higher level than the representation of a whole object. Importantly, this account is compatible with the fact that some patients with Charles Bonnet Syndrome were not aware of their hallucinations' unreality until they were corrected because "the ordinary appearance of their hallucinations, fitting well in the surroundings, made it very difficult or even impossible to discriminate real from unreal (Teunisse, Zitman, Cruysberg, Hoefnagels, & Verbeek, 1996, p. 796). This also suggests that felt reality is connected to the coherence of the objects with a scene.

3.6. *Conclusion*

In this chapter, I solved two problems related to perceptual presence: the problem of representation and the problem of subjective veridicality. The existing literature on perceptual presence conflates these problems, causing confusion in the debate. The first problem involved elucidation of the kinds of representation at play in perceptual presence and examination of the perceptuality of perceptual presence. I argued that there are three kinds of representation at play: representation of visible parts, representation of unseen parts, and representation as of the whole object. I argued that all these representations are perceptual under the predictive processing framework.

The second problem I tackled was how to account for the difference between experiencing something real and something unreal. I compared existing accounts (Seth's and Hohwy's) and developed my own account. I also argued that action elements play an important role, but that their role should be understood as a higher-level representation rather than as a counterfactual.

I focused almost exclusively on visual cases, but my argument can easily be extended to other sensory modalities such as tactile sensation (as in the bottle case), and auditory sensation. Auditory cases are interesting enough to be worthy of further exploration, because they naturally lead to the problem of temporary presence. For example, you might hear the same tone as being part of one musical piece or another. Even though you only hear a single tone in a given moment, you still hear the music as a whole. That is the topic for Chapter 5.

4. *The second P: Poverty*

4.1. *Introduction*

The second P to be elucidated is Poverty. Poverty is related to the so- called "the *Grand Illusion Hypothesis*". Intuitively, it seems to us that we enjoy a richly detailed visual experience across our entire visual field. Suppose you are at a botanical garden in spring to refresh yourself. You will enjoy a multitude of experiences of colors and shapes—flowers of different colors are blooming, the leaves of trees and grasses are exhibiting different shades of green, the trees are presenting their unique shapes. In having these experiences, you might think all the trees and flowers—including the ones in the periphery of your visual field—are simultaneously experienced with their specific shapes and colors, as in a photograph. However, the grand illusion hypothesis contradicts this intuition, claiming that our visual experience is actually much poorer than it is assumed to be. According to this hypothesis, we have rich experiences only of objects to which we are attending; the intuitive impression that we have a rich experience across our whole visual field is illusion. However, Jonathan Cohen (2002) argues, contrary to what the Grand Illusion Hypothesis assumes, that naïve people do not have introspective beliefs about the richness of their experience. Thus, the debate about the poverty of our perceptual experience has two parts: one about vision itself and the other about introspection.

In this chapter, I will discuss both. I will argue that even though people do have some form of belief about their visual experience, they do not believe that their visual experiences are like snapshots. However, I will argue that despite this, there is a problem of poverty that needs to be answered. I consider two accounts to solve this problem, and show that of these, the predictive processing account is simpler and does not suffer shortcomings of higher order accounts.

The Grand Illusion Hypothesis is motivated by findings from cognitive science, specifically work on change blindness and inattentional blindness (Mack & Rock, 1998; Simons, 2000). Empirical findings demonstrate that people are not very good at noticing things or events that are not attended or not foveated,[29] even when they occupy large

[29] Fovea is a central pit in the retina responsible for sharp clear central vision. To foveate is to move eyes so that one can see an object clearly with fovea.

areas of one’s visual field. This seems to show that we experience only those things and features which are attended to or foveated. Our intuitive belief about the nature of conscious visual perception is an illusion—that is the idea of the grand illusion (Noë & O’Regan, 2000).

The idea of the Grand Illusion is comprised by two components. The first component, which I will call the *Introspective Component*, concerns the introspective belief of perceivers’ that our visual experience is rich across our whole visual field. According to the grand illusion hypothesis, naïve subjects believe that unattended objects presented in our peripheral vision are perceived in full detail. The second component, which I will call the *Vision Component*, concerns the actual nature of visual phenomenology rather than people’s beliefs. If the Grand Illusion hypothesis is correct, the nature of visual phenomenology revealed by empirical studies, directly contradicts the introspective component: we can experience things and features only when they are attended to.

In the first half of this chapter, I argue that the Grand Illusion Hypothesis is false. Doubt can be cast on both components. Regarding the Introspective Component, even though people do seem to overrate the clarity of their visual experiences, there is nevertheless reason to think that our peripheral visual is not widely seen as being as richly detailed as the areas of focus. Attributing the widespread introspective belief that our visual field is uniformly detailed with crisp colors and clear shapes seems too strong a claim to assume without evidence, and this evidence is not provided. Similarly, regarding the Vision Component, it is not clear that the empirical evidence the supporters of the grand illusion hypothesis rely on unequivocally shows the poverty of our visual experience. Thus the upshot of the first half of this chapter will be that the grand illusion hypothesis itself is a kind of “illusion” (Cohen, 2012).

Despite this, I will argue that there is nevertheless an important poverty of perceptual experience, which I will call *informational poverty*. I will show, based on other kinds of evidence about our visual systems, that this is another rendition of the inverse problem introduced in Chapter 1, which is about how the brain represents the sensory cause given the many to many relationship between cause and sensory stimulation. The latter half of this chapter will be dedicated to the solution of this problem. I will argue that there is a way to overcome this poverty and that our subjective visual phenomenology can be much richer than suggested.

I will discuss the solution provided by higher-order theorists of consciousness (Lau & Brown, forthcoming; Lau & Rosenthal, 2011), and provide my own account based on predictive processing. I argue that my account does greater justice to visual phenomenology and hence has more explanatory power. Additionally, it is more parsimonious than the higher-order account and does not suffer from the problems that higher-order account faces. As it will turn out, the problem of poverty is similar to that of perceptual presence: both problems regard how to represent the world given insufficient sensory stimulation. Thus the solution is similar, albeit with some important differences. The other three problems set out in Chapter 1 will also be discussed: the description problem, the perceptuality problem, and the reason problem.

In this chapter, I will only talk about visual modality, but the same problem can arise for all sensory modalities. Furthermore, the poverty problem has implications for our stream of consciousness, which is often conceived as containing multitude of experiences spanning over multiple modalities (Dainton, 2006; James, 1890/2013). However, if we experience only what we attend do, our stream of consciousness can contain only a few items. I address this issue at the end of the chapter, after I give an account of visual consciousness.

4.2. *What is the Grand Illusion Hypothesis?*

In this section, I start by clarifying what the Grand Illusion is. After that, I will proceed to an assessment of the hypothesis.

I will begin with a review of the empirical findings related to the Grand Illusion Hypothesis. The Grand Illusion Hypothesis is motivated by phenomena related to two factors in conscious vision: attention and peripheral vision. Even though they are distinct factors whose effects should be evaluated separately, they are often treated jointly, partly because peripheral vision does not usually attract attention. However, these two factors are dissociable in practice: you can attend to the peripheral areas of your visual field. Thus in different experiments, these two factors can play out differently. In reviewing the experimental results, I will pay attention to how they work in each setting and will argue that the evidence obtained from peripheral vision is stronger than the evidence from inattention.

Let us start from reviewing evidence related to attention (or inattention). This group of evidence is taken to show that there is no conscious perception outside of attention.

The Grand Illusion Hypothesis was inspired by two surprising psychological phenomena: change blindness and inattentional blindness. In typical change blindness experiments, a scene is presented to a subject, followed by a blank screen or brief appearance of high contrast shapes scattered over the image, followed band after that a scene identical to the first one, except for one change. The task for test subjects is to detect the change. However, the task turns out to be far more difficult than expected by subjects. Furthermore, this holds even when the change has taken place across a relatively large area[30]. This change blindness is surprising, because once someone has detected the change, it becomes difficult to not see it. In a more realistic and perhaps more illustrative example, a pedestrian is asked directions from an experimenter (Simons & Levin, 1998). Their conversation is interrupted by some people carrying a door. Behind the door is another experimenter, who replaces the original experimenter and continues the conversation with the pedestrian. More than half of the pedestrians did not notice that they were talking to a different person!

Another phenomenon that inspired the Grand Illusion Hypothesis is inattentional blindness. This is a phenomenon where a subject tends to miss a clearly visible big change when they engage in an attention demanding tasks. A famous example is the gorilla experiment (Simons & Chabris, 1999).[31] A subject is asked to watch a game of basketball and count the number of passes. During the task, a gorilla emerges from one side of screen, walks across, dances, and disappears off the opposite side of the screen. Even though the gorilla is there for quite a long time, draws attention to himself, and also walks past the center of the visual field, almost half of the subjects didn't notice its presence. In both change blindness and inattentional blindness, subjects' failure of detection is due to lack of attention rather than foveation as in change blindness and inattentional blindness experiments, foveation was not constrained.

There is also recent work concerning conscious perception outside of attention. It has been claimed that at least some kinds of perception—for example, perception about gist (general summary) of a scene can be perceived without attention (Mack & Rock,

---

[30] Some demos can be found on Kevin J. O'Regan's website (http://nivea.psycho.univ-paris5.fr/ECS/ECS-CB.html).

[31] Demos of the experiment are readily obtainable, for example on Dan Simons' website (http://www.dansimons.com/videos.html).

1998). However, Michael Cohen (M. A. Cohen, Alvarez, & Nakayama, 2011; M. A. Cohen, Cavanagh, Chun, & Nakayama, 2012) argues, through the use of dual task experiments, that there is actually no perception outside attention. In the dual task paradigm, a subject is instructed to focus on one task, during which time he or she is presented with images irrelevant to the task at hand. The subject is then questioned about the images presented. Answering questions about the images is actually the task of interest, and measures the level to which a subject can detect images without attention. Cohen shows that subjects fail to detect gist representation presented in their center of visual field when the task they are instructed to do is more attention demanding than the ones used in the previous experiments (e.g. tracking of objects through time). Cohen claims this shows attention is a requirement for conscious perception; there is no conscious perception without attention.

The evidence presented so far is taken to suggest that there is no perception outside of attention. But other evidence related to foveation is also supposed to show the poverty of our peripheral vision. In this case, the alleged conclusion is different from that suggested by the (in)attention experiments. Dennett (1991) provides one illustration of the limits of perceptual richness, one you can try for yourself. Draw a card without looking it, and keep looking straight ahead throughout. Keeping the card at the edge of your visual field, hold the card at arm's length. Then move it towards the center of your visual field up to the point you can determine the color of the card. To the surprise of most people, you cannot identify the color of the card until it comes almost to the very center! This is due to the structure of our retina. This experiment dissociates attention and foveation—you keep looking straight so you are not foveating but attention is nevertheless directed at the card.[32]

Supporters of the Grand Illusion Hypothesis use such evidence to support their conclusions. For example, Noë and O'Regan say:

> It does not seem to us as if we only see that to which we attend. It seems to us, rather, as if we are perceptually aware of the densely detailed, stable and persistent environment around us. But since we do not attend to all that detail, at least not all at once, then it would seem to follow that perceptual

[32] Another illustration is provided by Schwitzgebel (2008).

> consciousness—that feeling of awareness of all the detail—is misguided. (Noë & O'Regan, 2000, p. 2)

I think the most neutral way to interpret their claim is to interpret it as a thesis about content of perception. The term 'content' is ambiguous and is often associated with the idea that experience represents–I also introduced phenomenal representational content in Chapter 2. However, since some supporters of the Grand Illusion Hypothesis, such as Noë, are anti-representationalist, it makes sense to begin from a less loaded version of the notion of content. I will therefore start with Bill Fish's usage of content:

> We must be aware that there is a fairly innocent spatial understanding of the term "content". For instance, the "contents" of my pocket are, at present, some coins, keys, a cellphone, and lint. But the fact that my pocket has contents, in this sense, does not mean that it has correctness conditions or that it is potentially true or false. So we need to be aware that, sometimes, when people talk about the "content of perception" it is used as a way of referring to what is in our experience or what is perceived. (Fish, 2010, p. 9)

Thus, under this usage of content, the content of conscious perception just refers to what is experienced. The thesis of interest here concerns the range of things in the world that we can (visually) experience, along with their properties. When we experience things in the world they are in our experience and vice versa. With this notion, it is natural to think that richness or poverty is richness or poverty *of the content of visual experience.* Let us unpack Noë and O'Regan's claim using this conception of richness and poverty. The Grand Illusion Hypothesis seems to have this form:

> Grand Illusion Hypothesis
> Naïve people believe that the content of visual experience is rich (Introspective Component). However, content of visual experience is not rich (Vision Component).

Now, what it is for the content of visual experience to be rich needs to be cashed out. Noë and O'Regan (2000), quoted above, stated that we only see what we attend to, but that we believe that and our visual world is uniformly rich. Thus, one way to formulate *Richness* is the following:

> Richness
> The contents of visual perception are rich when and only when a subject experiences all visible properties of all visually experienced objects at the same fine-gained level.

Two closely related claims are *Representational Richness* and the *Representational Grand Illusion Hypothesis*. Richness and the Grand Illusion Hypothesis themselves do not involve representational content, they only involve content in Fish's sense. However, the standard position both in philosophy and cognitive science today, with some exceptions among anti- representationalists such as Noë, is that perceptual experience represents. Perceptual experiences are about the things in the world and perceptual experiences obtain intentionality in virtue of representing those things. This assumption leads to representationalism:

> Representationalism
> When we experience things and their properties in the world in a certain way, we represent them in that way by virtue of their phenomenology.

We also get representational richness:

> Representational Richness
> The representational content of visual perception is rich when and only when a subject has a conscious representation of all visible properties of all visually experienced objects in the same grain of fineness.

These definitions echo a statement by Ronald Rensink, a proponent of the Grand Illusion Hypothesis:

> One of our most compelling impressions as observers is that we are surrounded by a coherent, richly detailed world where everything is present

simultaneously. Although our environment is certainly this way, this impression is so compelling that we tend to believe these properties true of our representations as well—that is, we believe that somewhere in our brain is a stable and detailed representation of the stable and detailed world around us. (Rensink, 2000, p. 17)

The corresponding version of the Grand Illusion Hypothesis would be:

Representational Grand Illusion Hypothesis
Naïve people believe that representational content of their visual experience is rich (Introspective Component). However, representational content of visual experience is not rich (Vision Component).

In the remainder of this section, I will examine if any of these versions of the Grand Illusion Hypothesis holds. To begin with, we will see if the Vision Component actually follows from the empirical evidence reviewed in this section. It turns out that it is not at all obvious, because proponents of the Grand Illusion commit inferential errors with regards to the Vision Component.

What change blindness and inattentional blindness establish is that people sometimes cannot report changes of properties or objects in their visual fields. But this alone is insufficient to establish that these properties or objects are not experienced at all. It seems entirely possible that something can be visually experienced but that its content cannot be reported because it was not attended to. There is a famous debate regarding reportability and perceptual consciousness (Block, 1995, 2007), and there are two senses of reportability. On a stronger reading, a representation is reportable only when it *is* attended to, but on a weaker reading, a representation is reportable when it *could* be attended to. Block (1995, 2007) famously argues that representations reportable only in the second sense are also conscious (he also adopts stronger sense of reportability, claiming that unreportable states can also be conscious). However, others, such as Prinz (2012), think attention is necessary for consciousness, and thus only representations reportable on the stronger reading are conscious. In the context of change blindness and inattentional blindness, to conclude that states that are not reportable due to inattention are not conscious is to commit to the thesis that states have to be reportable in the stronger sense to be conscious. However, the debate has not been concluded yet thus this reasoning is hasty.

This point is recognized by Dan Simons and Ronald Rensink, two of the main researchers in the field:

> Unfortunately, these important advances have been clouded by the drawing of several overly strong conclusions. […] the existence of change blindness does not on its own necessitate sparse representations – it could occur even with fairly detailed and complete visual representations of a scene. (Simons & Rensink, 2005, p. 17)

By sparse representations, they mean representations of objects with fewer details, or of properties of objects in a coarse-grained description. Here, they do not say there *are* detailed conscious representations of those objects/properties. They simply point out that change blindness and inattentional blindness are compatible with visual experience rich in content, with some of that content not being reportable due to limitations of attentional resources.

What about the Introspective Component? Some philosophers, such as Jonathan Cohen (2002), raise doubts about the idea that lay people hold the introspective beliefs of the Introspective Component. His doubts stem from the nature of the concept of representations. The concept of representations is not the kind of the concept ordinary perceivers has—representations are theoretical entities in cognitive sciences (and philosophy). For example, grasping mental representation in the case of vision involves grasping causal dependency between external objects and brain states and the possibility of misrepresentation. Ordinary perceivers wouldn't care about these things, instead thinking they experience external things directly (hence, Naïve realism). Because of this, Jonathan Cohen thinks ordinary perceivers cannot have introspective beliefs about representations.

However, this objection can be answered as follows. First, Cohen's objection does not apply to Richness (nor to the corresponding version of the Grand Illusion Hypothesis). People can have beliefs about the content of their visual experiences; this only amounts to saying that ordinary perceivers have some beliefs about what they can experience at a given time. Naïve Richness does not make any reference to the concept of representation and is thus immune to the objection Second, ordinary perceivers do not need to have the *concept* of representation in order to have beliefs *about* representation. This is another instance of the familiar distinction between sense and meaning (Frege, 1948). For example, most people who observed Superman flying would form a belief about Superman. However, this same belief is also about Clark Kent, even though they do not recognize the identity relationship, and so do not believe Clark Kent can fly. In the same way, people might only have concepts for what they can experience, yet might unknowingly form beliefs about representation. People can have beliefs about what objects in the world they visually experience. Thus they can have beliefs about representational contents, because at least part of what they experience comprises representational contents if perceptual states are representations. For instance, a person might experience a pink elephant in front of her—it seems that she directly experiences a pink elephant in the world but it is actually an intentional object of her representation.

Another objection to the Introspective Component comes from Bayne and Spener (2010). They state that the Introspective Component is not ubiquitously held among perceivers and that some people have intuitions to the contrary—some people believe the phenomenal content of their visual experience is rather impoverished. They refer to Zimmerman (1989) in which he claims that his *personal experience* tells him the contents of experience are limited to what he is attending. Even though the comment is a personal one and there is no trouble in finding scholars endorsing Richness, Zimmerman's claim cannot simply be dismissed; it is a serious claim by a thoughtful scientist. Thus it seems that although most people have unreliable introspection and erroneously believe in *Richness*, this is not ubiquitous. Schwitzgebel (2008) is a skeptic about introspection, and he writes about this topic of richness/poverty of visual consciousness. In his paper, Schwitzgebel argues that although people intuitively believe Richness, they can discover the poverty of periphery after receiving some "training." One mechanism for this is learning to dissociate attention from fixation. Normally, when we attempt to pay attention to some item presented in the periphery, we also fixate on that item. But that is not mandatory. We can actually separate attention and fixation, attending to things we do not fixate on. To illustrate, suppose, while at a café, that you spot someone you wish to avoid. You can focus your attention at them while keeping your face and eyes directed at something different—your book or your friend, perhaps. By learning this skill, perceivers can now attentively investigate what their peripheral vision is like, and discover how poor the content of their peripheral vision is! Schwitzgebel is skeptical about the value of introspection in many fields, but this example indicates that people *can* develop different introspective beliefs through training in the case of visual richness or poverty[33].

This suggests that through training, Poverty can be revealed to subjects, causing them to let go of Richness. Nevertheless, for naïve perceivers the Grand Illusion Hypothesis is sustained. Poverty is revealed to perceivers only when they learn to dissociate attention from foveation. I call this *Revealed Poverty*, and it becomes an explanandum for a satisfactory account of poverty. I will later argue that *Precision Expectation* can explain Revealed Poverty, but first, a definition:

> Revealed Poverty
> If a perceiver attends to their parafoveal area, the subject can discover that there are not many details in their peripheral vision.

Moreover, there is also reason to think subjects don't believe in literally picture-like uniformly detailed representation. As Alva Noë notes about our daily banal actions:

> Notice that we are not surprised or in any way taken aback by our need to move eyes and head to get better glimpses of what is around us. […] The fact that we are not surprised by our lack of immediate possession of detailed information about the environment shows that we don't take ourselves to have all that information in consciousness all at once. (Noë, 2004, p. 58)

When you look at an electronic timetable at a train station, you turn your head and eyes straight at that. You would not try to read the train schedule while looking at it in your peripheral vision. This shows that on some level you, even if you are not aware of this recognition, know that you can see things better when you attend to them, and that

[33] Schwitzgebel (2008) is more pessimistic about other cases of introspection of conscious experiences such as the nature of emotional experiences and existence of imageless thought.

you see more clearly in the center of your vision. This is not to say that we regard peripheral visual field as total blank or unattended object as nonexistent. Rather, what this shows is that perceivers implicitly understand that objects in or peripheral vision, or objects not attended to, are experienced with fewer details. In other words, perceivers implicitly understand Poverty to some degree. I call intuitive understanding *Moderate Poverty*, which is another explanandum an account of poverty must explain.

> Moderate Poverty
> Recognition of the fact that peripheral unattended areas in the visual field have less details than the foveated attended area is usually reflected in behavior.

In this section, I formulated two theses which together make up the Grand Illusion Hypothesis: the Vision Component and the Introspective Component. I argued that change blindness and inattention blindness fall short of proving the claim of the Vision Component. However, regarding the Introspective Component, I argued that, apart from cases where attention is dissociated from foveation, we do have the kind of introspection relevant to that component. This means that if Poverty is shown to be true, then the Grand Illusion Hypothesis is also true. I also provided two explananda that an account of poverty has to explain: Revealed and Moderate Porverty. In the next section, I will turn to what I take to be stronger evidence for Poverty (which does not allow for the possibility of unreported representations of richly detailed contents) and will argue that Poverty is not entailed by the evidence yet.

4.3. *The physiological evidence*

In the previous section, I argued that change blindness and intentional blindness fall short of proving the Vision Component. However, I will now argue that physiological evidence does a better job.

According to the physiological evidence, due to the structure of the human retina, peripheral vision does not have fine-grained information about color or shape There are two kinds of photoreceptors in the retina: rods and cones. Cone cells contribute to higher visual acuity than rod cells thanks to their one-to-one connection with a ganglion cell and the optic nerve and cone cells are also dedicated to color perception because the three different kinds of cone cells are activated by light of three different kinds of wavelengths. These cells are mostly operative under sunlight as they have lower sensitivity to light and take much energy to be activated. Conversely, rod cells can only provide lower acuity (multiple rod cells are connected to one ganglion cell) and they do not mediate color perception. However, they play a large role in scotopic vision (vision in the dark) as they have higher sensitivity to light. This is why we cannot see much color in the dark but can still perceive shapes. The distribution of the two kinds of photoreceptors tells us interesting things about visual phenomenology. Even though rod cells are distributed throughout the retina, cone cells exist almost exclusively in the foveal area. Moreover, the area in our visual field that corresponds to fovea covers only the area of a thumb tip at arm's length. If the content of visual experience supervenes on informational content deriving from sensory causes, then we can experience crisp colors and clear shapes only in a tiny area of our visual field!

Furthermore, the existence of a blind spot in the retina is well known. There are no photoreceptors in the area where the optic nerve passes through the retina. Saccadic suppression adds another piece of evidence. The eyes saccade to a different point every 200–300 milliseconds, and during the saccadic movements, the visual signal is suppressed—the signals received during saccades are not processed.

These pieces of evidence together suggest that our visual experience is something like this: only a tiny area of the visual field the size of a thumbnail at the arm's length has specific shapes and fine-grained colors, there is a small "hole" in the visual field, and vision completely blacks out every 200 milliseconds. Yet our experience does not match these facts.

These points provide stronger evidence for the Poverty claim. Change blindness and inattention blindness both focus on the "downstream" side of cognitive processes; it could be claimed that objects and their properties are being represented but just not used by the reporting system. However, the evidence reviewed here rejects this possibility because it concerns the "upstream side" of the processes. That is, it is not possible that there is unreported representation because the evidence shows that the sensory organs do not receive sufficient information. Faced with this evidence, acceptance of. Poverty seems required, as there is simply no information to be perceived

Yet this need not be true. The inference to Poverty from the "upstream" evidence is based on the assumption that the contents of perception have to be derived from information received by sensory organs. However, this is not necessary. The brain might "hallucinate" or otherwise fill in the contents of perception; the contents of perception might be yielded in a top-down fashion to make up the poverty of processed input. Two accounts that posit such a top-down mechanism are the higher-order account and the PP account, both of which will be examined shortly. Both accounts aim at overcoming this problem: Poverty of Information. This *Poverty of Information* together with *Revealed Poverty* and *Moderate Poverty* are the explananda for the accounts.

> *Poverty of Information*
> There is less sensory information arriving to the retina than it seems in visual phenomenology

As mentioned above, poverty of information is surmountable by top-down mechanisms that can supplement the lack of sensory information.

These two explananda are addressed in Section 4.5. I will start with the higher-order account and see how it accounts for these explananda.

4.4. *The higher-order account*

The first solution is provided by higher-order representational theories of consciousness (Lau & Rosenthal, 2011; Lau & Brown, forthcoming). Higher-order representational theories of consciousness are one set of promising theories of consciousness that have both philosophical and scientific defenders[34]. They come in many varieties, but the core claim shared among all versions is called the Transitivity Principle:

> Transitivity Principle
> A mental state is conscious only if the subject is conscious of the mental state.

Thus these theories explain consciousness of a state in terms of subjective awareness of

[34] Supporters in philosophy include Carruthers (2003), Rosenthal (1986, 1997, 2002, 2005), Lycan (1996).Supporters in science include: Lau (2007) and Rolls (2004).

that state. In the higher-order tradition, this awareness of the state is representational;[35] by virtue of having a representation about a mental state, a subject is aware of the mental state. The higher-order representational state makes the lower target state conscious, while the higher-order state itself usually remains unconscious unless there is an even higher-order state whose target is the higher-order state. The nature of the representational state that is responsible for subjective awareness diverges among theorists. Some argue that is thought (Rosenthal, 1986, 1997, 2002; Carruthers, 2003), some argue that is perception (Lycan, 1996), others argue that is a metacognitive judgment about the reliability of inner signals (Lau, 2007).[36] Another relevant characteristic of these theories are that all hold that the contents of conscious experience are the contents of the higher-order representation. This is because, for higher-order theorists, the contents of conscious experience are how we are conscious of the (lower-order) mental state.

Some higher-order theorists argue that the meta-representational structure and the explanation of phenomenal contents by virtue of the contents of higher-order representation, enable higher-order theories to provide a better solution to the problem of the poverty than lower-order theories can (Lau & Rosenthal, 2011; Lau & Brown, forthcoming).They argue that because early informational sensory processing is missing, a first-order theory has to accept none but the very impoverished picture of visual phenomenology. However, if a version of higher-order theory is correct, the mechanism posited in the higher-order theories allows for richer picture of visual phenomenology. The basic idea is that since the contents of conscious experience are the contents of the higher-order mechanism, they need not correspond to the information the brain carries. The higher-order mechanism can *misrepresent* the contents of the first-order state and

[35] An important difference within higher-order theories is that in some theories just the disposition to form a higher-order representation is sufficient for consciousness, while in others there needs be an actual higher-order representation.

[36] Note that this principle only specifies a sufficient condition for a mental state to be conscious. We can be aware of our mental states via conscious inference (perhaps by observing behaviors by yourself) or being taught by someone else. But these ways of being aware of one's mental state doesn't give rise to conscious mental state. Thus, when a mental state is conscious, a subject is aware of that state in a way that does not involve conscious inference.

create rich visual phenomenology. Indeed, Rosenthal (2002) argues in the following way, taking Andy Warhol's famous painting of multiple pictures of Marilyn Monroe arranged in a grid (Marylin Diptych) as an example:

> Other striking examples occur in connection with our perceptual sensations. […] parafoveal vision can produce only low-resolution sensations of most of the Marilyns in Warhol's famous painting, but we are aware of them all as clear and focused. What it's like for us is a function not of the character of our sensations, but of how we're conscious of those sensations. (Rosenthal, 2002, p. 415)

However, what contents does a higher-order representation have given the absence of first-order informational contents? How is it possible? One possible way is that a higher-order representation creates the contents in a top-down fashion. Part of the contents might come from information from past saccades or from general contextual knowledge. A higher-order representation might even fabricate some of the contents.

Hakwan Lau (2007, 2008) has a more neurological account of how and why this happens. According to his theory, based on signal detection theory, conscious perception takes place when the strength of a signal exceeds a criterion. To set the criterion, the brain has to estimate the level of noise in the brain (what its baseline level of activity is). Lau argues that this is a higher-order mechanism because the mechanism has a representation about the first-order state rather than about objects in the world. Lau is also able to use his theory in combination with empirical evidence to explains how a higher-order representation makes up the contents in unattended areas (Lau & Rosenthal, 2011; Lau & Brown, forthcoming).

It is known that attention improves the noise-to-signal ratio of the internal signal (Briggs, Mangun, & Usrey, 2013). Thus, under inattention, the signal fluctuates a lot because of the increased noise level. However, because of this fluctuation, the internal signal happens to satisfy the criterion for conscious perception just discussed. This could be prevented if the human beings could employ different criteria for unattended signals; setting a higher criterion for highly variable signals so that fickle criterion passing would not happen. However, there is strong empirical evidence that human beings can use only one criterion. The upshot of this is that we can predict conscious perception despite its low quality signals, albeit misleadingly. This is supported by the work by Rahnev et al. (2011), who show that inattention leads to liberal detection bias and higher visibility ratings.

This account raises some questions. First of all, higher-order representation is supposed to be located in prefrontal area, which is thought to have lower informational capacity than occipital areas. So wow can representations in prefrontal areas sustain rich representations? Second, if the account just given is correct, the higher-order representations *mis*represent the lower-order state all the time even though the embedded first-order content (I think that I am in the mental state that…) is veridical most of time. This is mysterious. All things being equal, misrepresenting something is not a good thing for a creature. How bad it is for a creature to misrepresent its own mental state is not as clear as the external case. Nevertheless, if Lau's account (or one like his) is correct, this systematic misrepresentation requires a biological explanation. Thirdly, it is questionable if Lau's theory is higher-order in the philosophically important sense. It is true that the account posits a mechanism which supervises internal signals, but this is not sufficient for a representation to be higher-order. A representation is the end product of the mechanism, and its contents are determined by how it is used by other mechanisms. The representation that the alleged "higher-order" mechanism produces seems entirely first-order in this respect. According to Lau's theory, the end product here is a conscious visual percept, and assessment of the internal signal is performed in order to decide whether the internal signal is a representation of an external object. It could be argued that the metacognitive assessment of the clarity of experience or confidence in perception is included in the content of representation, but this seems problematic. If metacognitive assessment is included in the representation, it cannot account for all kinds of conscious visual percepts, as most of the contents of our conscious visual experiences do not include metacognitive elements. The contents of our daily visual experience seem to be of the form "There is a tree in front of my red car" or "A cat is sleeping on a mat".

To conclude this section, let us see how this higher-order account fares with the explananda considered in the end of last section. I argued that *Informational Poverty* can be overcome by a top-down mechanism that can supplement sensory information. On this higher-order account, metarepresentational mechanism is this device. However, this came with a price: the metarepresentation is about one's mental state, thus higher-order representation misrepresents the poor sensory state. Conversely, the account based on predictive processing is an account that also makes use of some sort of hierarchical structures and top-down processes, yet does not suffer from the problems of the higher-order account because it does not necessitate any metarepresentational mechanism.

What about *Moderate Poverty*? The Higher-order account does not directly address this point but it has some resources. In some cases, higher-order theorists seem to think that richness of visual phenomenology is an illusion. In discussing interpretations of Sperling's 1960 experiment, Rosenthal (2007) and Brown (2014) both endorse so-called partial awareness hypothesis (Kouider, De Gardelle, Sackur, & Dupoux, 2010).[37] In the experiment, subjects are required to report on alphanumeric characters arranged in a grid-like way and presented only briefly. Subjects can only report a fraction of the characters, yet subjects report that they have seen all of them. One straightforward interpretation of the subjects' reports is that they have seen those characters with their own unique identity and full details[38]. However, supporters of the partial awareness hypothesis take subjects to mean they have seen all the characters but only as *characters in general*; they have seen something like letters or numbers, but they are not sure which specific ones they are. This way of interpreting the result suggests that subjects do see something in their peripheral vision, but this is less detailed than attended or foveated part of visual field. This seems a good answer for *Moderate Poverty*. However, the answer is open for other accounts, including the predictive processing account.

Regarding *Revealed Poverty*, the higher-order account does not provide any account at all. This is not to say the higher-order account is wrong because of this, but it shows that higher-order account is not sufficient for this aspect of the problem of poverty. As we can see, the higher-order account only provides an explanation for *Informational Poverty*, and that explanation is still problematic. In the next section, I will argue that the predictive processing account can better account for all of these factors.

---

[37] Rick Grush (2007) also provides a similar account.

[38] This is how Block (1995, 2007) takes the result of experiment.

4.5. *A predictive processing account*

In the previous section, we examined an account answering the problem of poverty proposed by higher-order theorists of consciousness. Even though it succeeded in explaining Informational Poverty to some extent, it has some problems, and was unable to explain some aspects of Poverty.

I will argue that predictive processing is best-suited to account for the problem of poverty because the problem of *Informational Poverty* can be considered another rendition of the famous "inverse problem": the brain is in the business of representing the cause of the sensory stimulation but the sensory stimulation the sensory organs receive *underdetermines* the cause. Underdetermination takes place because any given sensory stimulation is compatible with many possible causes. This is exactly what we have here. The periphery receives only scarce information and is therefore compatible with many possible worldly causes. Predictive processing provides a solution to the inverse problem, and can therefore provide an account of poverty.

In Chapter 1, I reviewed elements in the predictive processing account are useful to explain various perceptual phenomena. In this chapter, the tools to marshal to explain the problem of poverty are: *Prediction Error Minimization*, *Precision Expectation*, and *Hierarchy*. Prediction Error Minimization equipped with Hierarchy and Precision Expectation can jointly account for the explananda above. The Poverty of Information is overcome by the hierarchical processing.

As it has priors, perceptual content not strongly constrained by sensory stimulation on the sensory organs. How does this work? The higher-level models in the hierarchy play a pivotal role in providing priors. They deal with a longer spatiotemporal timescale and provide contextual information for the lower-level models. For example, suppose you are looking at a bustling street scene in Tokyo (Figure 9). When visually experiencing this, one higher-level model might represent that it is a cool urban scene. Then this provides a contextual prior for the lower-level models. It constrains the lower-level model, telling it what kind of objects are typically included in this kind of scene, and what kinds of colors and shapes to expect.

The basic story to be told here is similar to the one I used to explain perceptual presence in Chapter 3, pending one important difference to be discussed shortly. In Chapter 3, I invoked three different representations: two lower level models (representation of seen parts and representation of unseen parts), and one higher level model (representation of the whole object) (Figure 3). Representation of seen parts roughly corresponds to representation of attended and foveated objects and their properties; representation of the whole object roughly corresponds to representation of the scene category; and representation of unseen parts roughly corresponds to representation of unattended objects and their properties.

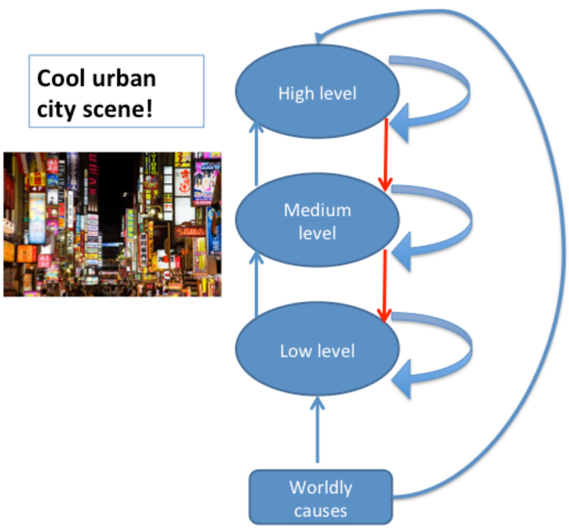


Figure 9 Perceiving a cool urban scene in Tokyo

The important difference here is how precision expectation works. When the precision of prediction error or sensory stimulation is estimated to be low, the model is not revised based on sensory stimulation. Rather, it relies on priors that are furnished by top-down prediction from the higher-level models. In the case of the poverty here, the brain learns that sensory stimulation from the periphery contains a lot of noise and so ignores it. Empirical evidence supports this interpretation. According to the evidence, when things are perceived peripherally, the perception is influenced by information obtained from previous saccades (Herwig, Weiß, & Schneider, 2015). In such cases, where there is insufficient information provided, higher-level models could come from past experience, or it could be that there is a 'quick and dirty' route to the top of the hierarchy that sends the estimation of the scene category (Hochstein & Ahissar, 2002).

This working of precision expectation is radically different to the case of the representation of invisible parts in perceptual presence. In the case of the representation of invisible parts, I argued that associated phenomenology is weak because the top-down prediction from the higher-level model does not meet bottom-up prediction error. This also applies here. The higher-level model of the urban scene predicts matching objects at the lower-level, but in this case, there is no sensory stimulation received from those objects. Nevertheless, we can still experience those objects to some extent because the prediction error at that level is estimated to be low and so priors play a larger role in perceptual inference. The outcome is that the representation of the kinds of objects that tend to exist in urban scene is assigned higher probabilities. In this way, Informational Poverty is overcome: despite the lack of sensory stimulation, objects and their properties are experienced through relying on priors.

An explanation of *Moderate Poverty* requires a more detailed analysis of the subject matter. One explanation for *Moderate Poverty* is that there is only a representation of general information (such as the presence of alphanumeric characters) without detailed information being present. It seems true, for example, that we experience all the characters as characters, but they are not experienced without color or shape The details of unattended objects may not be as fine-grained as attended ones, but nevertheless they are experienced as having color and shape. The entity being a character concerns object category, but this is different from the color or shape the entity has.

To elucidate this, I would like to propose two axes of analysis: the level of models, and the degree of the determinacy of properties. Levels are the levels introduced in the explanation of *Hierarchy*: One property is higher in level than another if the property deals with a longer spatiotemporal timescale. Determinacy concerns the fine-grainedness of the property. For example, red is more determinate than colored; scarlett is more determinate than red. Being red and being scarlet belong to the same category, but one property is more determinate than other.

In this situation, representations of an object's category can be regarded as a higher-level model than the models about color and shape. This is because representations of a category have to do with longer spatiotemporal scale; the object category concerns the whole object, while color and shape is can apply. Moreover, even though those models belong to different levels, there is an important relationship between them. As described previously, higher-level models predict and constrain lower-level models. Therefore, we can experience some color or shape even in the periphery thanks to predictions from the higher-level models. However, they have to remain indeterminate as there simply is not sufficiently detailed information available.

What remains is something like the shape-that-a-typical-buiding-in-Tokyo-has, for example. This account is remarkably different from the one given in higher-order account. In the higher-order account, only representation of the object category was explained. However, the predictive processing account can also account for some indeterminate experience of color and shape, which stays true to phenomenology.

What about *Revealed Poverty*? The predictive processing account holds that there is a real variation in experience between of naïve perceivers and perceivers who learn to attend to parafoveal areas. Precision Expectation can give a nice illustration here. I mentioned that attention is precision expectation. When attention is directed to a specific parafoveal area, the subject would expect high signal-to-noise ratio from that region. By doing this rather than ignoring the prediction error given from the parafoveal area, the brain takes in prediction error to revise its model. Consequently, this ends up showing indeterminacy of information in parafoveal area provides. Hence poverty is revealed.

Now, let us compare this predictive processing account with the higher-order account discussed before. First, the predictive processing account is a thoroughly first order account. It is quite right that precision expectation, which plays a key role in the account, is a second order statistical inference because it is an inference about the reliability of perceptual inference. However, the resultant operation gains control of error units (which decides how much prediction error will be taken in to the model), and no metarepresentation is formed beside the representation of worldly causes.

The predictive processing account does not face the difficulty of the higher-order account. The predictive processing account is not plagued with ubiquitous erroneous metarepresentations—according to higher-order account, metarepresentational states about sensory states routinely misrepresent—and the predictive processing account also presents a much more straightforward explanation. In the predictive processing framework, models, or representations, at adjacent levels work “in collaboration” to form the best hypothesis about the world; a prediction error that is not resolved at one level is passed one level up, and the higher-level model passes a prediction to the lower level. Only the first-order representation is dedicated to representing the world; the second-order representation is about the lower-order mental state, not about the world. Because of this, the second-order representation cannot compensate the poverty of the lower-order state, hence misrepresentation.

To conclude, the predictive processing account provides a straightforward solution to the problem of poverty and explains all the explananda. Moreover, the explanation is simpler and superior to the higher-order solution.

4.6. *Answering the description problem, the perceptuality problem, and the reason problem.*

Before closing this chapter, let us answer three other question raised in Chapter 1: the description problem, the perceptuality problem, and the reason problem. The answers to the description and reason problems can be found in the foregoing sections, but it is useful to explicitly address them here.

Answering the description problem

The problem of poverty presents a difficult topic for arriving at the right description of this aspect of perception. Half of this chapter was dedicated to drawing the right picture of the phenomenon, and it was rather complicated: we implicitly notice the poverty of visual experience but the information contained in experience is still more than the sensory stimulation contains. But this complicated picture was neatly explained by the predictive processing. This successful explanation by predictive processing in turn establishes the correctness of the description, through the “mutual constraint” I described in Chapter 1.

Answering the perceptuality problem

Given the predictive processing account, some might doubt its perceptual nature because of the involvement of higher-level representation. However, a response similar to that provided in the presence chapter works. High-level representations are also engaged in explaining away sensory stimulation and jointly form a percept with lower-level representations. Moreover, poverty is also solved at lower-levels by prediction from higher-levels. Therefore, I conclude that the richness of perception is solely a perceptual phenomenon.

Answering the reason problem

It is actually rather unnatural to ask why we experience a rich world, as the experiential richness originally stems from the richness of the world. This becomes surprising only when we discover the limitations of sensory stimulation. The question then becomes how we overcome this limitation. It turns out that the hierarchical prediction error minimization best explains this phenomenon.

4.7. *Conclusion*

In this chapter, we discussed the problem of the richness of visual consciousness. The nature of this phenomenon was more complicated than it seemed at the beginning, but predictive processing provided a nice illustration without the costs of endorsing higher-order theories of consciousness. The solution given in the predictive processing account was similar to that in the last chapter, apart from important difference regarding precision expectation. The main target of the account was visual consciousness, but we can apply the same story for the stream of consciousness in general: even though we can only attend to a few items, we can experience the whole stream of consciousness, thanks to higher level modes and the priors provided by them! The next three chapters will address rather different aspects of perceptual experience.

5. The third P: Present

5.1. *Introduction*

The focus of this chapter, and the third P to be explored, is Present: the temporal aspect of our mental life. Thus far, I have not discussed the temporal aspects of perception. The examples explored in previous chapters have been about the simple perception of an object or a scene without referring to any temporal aspects. However, the temporal aspect of consciousness is an integral part of our conscious life. First, perceptual experience is always about "now" and our conscious perceptual experiences "flag the present". This is also part of the core characteristics of perceptual experience: openness to the world (McDowell, 1994, p. 111). Namely, perceptual experience directly presents us with mind-independent objects as existing *here and now*. Perceptual experiences tell you about things that exist or events that are happening "now" to help you deal with imminent issues. In this sense my conscious perception of a red cup in front of me is different to the conscious episode of remembering a red cup or anticipating a red cup. Moreover, this experienced present, what William James calls the *specious present*, phenomenologically seems to have certain "breadth". For example, when you listen to the opening of Ludwig van Beethoven's *Symphony No. 5*, you hear the famous four-note "dit-dit-dit-dah" motif. The motif takes around two seconds to play and is repeated twice, but the whole motif seems experienced as a single unit.[39] In this way, we routinely talk about events which seem to be directly experienced despite having a certain duration. In other words, the specious present we experience is not like a volume-less mathematical point. As James (1890/2013, p. 609) argues, "[T]he practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time". However, this specious also carries a highly puzzling aspect. Even though the whole "dit-dit-dit" is experienced as happening now, it also is experienced as having a temporal order—the first dit is followed by the second dit which is followed by the third

[39] A two-second-long specious present might look long, but there is disagreement about the length. James (1890/2013) thinks it is six to twelve seconds, whereas Dainton (2006) thinks it is about half a second. However, what I need here is just a sequence of events that seem to comprise one specious present. So the actual duration of the example is not an essential feature.

dit. In other words, when the second dit is experienced, the first dit is experienced as having happened before the second dit while the third dit is experienced as happening after the second dit. How is it possible that one thing is experienced as happening now and also as happening before or after now? One of the main aims of this chapter is to give an adequate account of this puzzle of the specious present from the predictive processing perspective. The close examination of phenomenological aspects of the specious present reveals a nested, hierarchical structure, which matches the hierarchical structure posited in predictive processing.

Another question I will address is about the temporal properties of perceptual experiences *themselves*. We experience events that have a certain duration. However, strictly speaking, it is the *contents* of experience that have duration.[40] When we hear the opening motif of *Symphony No. 5*, we experience a motif that has duration. That is *apparent* duration: how the musical tones appear to you in your experience. This may or may not be same as the *actual* duration of the motifs. And it is yet another question if the duration of experience *itself* matches with either of these. Thus, we need to distinguish three items: the duration of events in the world, the apparent duration of events, and the duration of experience.

If you think perceptual experience is analogous to a movie, then all three seem to be the same: the duration of what you see in a part of a movie is the duration of the part of the movie played and also is the duration of the event being filmed. Thus, it might seem unmotivated to think they don't coincide. When you experience two opening motifs for four seconds, it seems natural to think the experience itself lasts for four seconds, or at least something close to it. The position holding that the duration of the contents of experience roughly coincides with the duration of experience itself is called *extensionalism* (Dainton, 2006, forthcoming).

However, on reflection, this position seems less plausible. I will discuss why in detail in the next section but, will touch on a general point here: it does not usually hold that experience has the properties that the objects of experience apparently have. For

[40] Skepticism about the very possibility of experiencing events with duration is also possible (Chadha, 2015). According to the skeptical view, the experience of events with duration is nothing but a collection of duration-less experiences, just as rapid exposure to static images creates the impression of motion. In this chapter, however, I assume that skepticism is false.

example, when we experience something red, it is not usually thought that our experience itself also has the property of being red (unless you are committed to sense datum theory). When you experience some pain on your left hand, it is not thought that your experience also has that spatial property. In the case of temporal consciousness, it is true that the situation is a bit different from these other cases. In the case of color experience, there is no reason to think experience possesses color properties. However, experience does possess temporal properties as the experience also exists in the world. Thus, it is not unreasonable to expect some mirroring relation between the temporal properties of experience and the apparent temporal properties of events. Still, it is not a given that there actually is such a relation.

In this chapter, I will summarize this debate finding that there is no conclusive argument for or against extensionalism. However, I will give some reasons to think that the main opponent of extensionalism, retentionalism*,* is preferable. Retentionalism is the position that experience itself is momentary, but can *represent* events as having duration. According to the position, the specious present is somehow "crammed into" a momentary present. I then argue that the predictive processing account presented here fits better with retentionalism than with exhtensionalism because the hierarchical phenomenological structure the predictive processing account calls for nullifies the motivation of extentionalism.

This chapter will proceed as follows. In the next two sections, I will give phenomenological explananda of the specious present. In Section 4, I will give a predictive processing account of those explananda. Section 5 will then give an overview of the debate between extensionalism and retentionalism. In Section 6, I will return to the predictive processing account, and discuss how it fares with extensionalism and retentionalism. In the last section, I answer the four questions raised in Chapter 1: the inverse problem, the description problem, the perceptuality problem, and the reason problem. As a reminder, the inverse problem is about how we can experience the specious present given that our sensory stimulation is limited; the description problem is about how to the describe phenomenology of the specious present aptly; the reason problem is about why we experience the specious present described in this chapter and the perceptuality problem is about whether our experience of specious present is perceptual or not.

In this chapter, I will limit myself to an analysis of the specious present of *perceptual experiences*. It is true that conscious episodes of imagining or recalling also exhibit a form of presentness. When you recall an episode of encountering a friend, your experience seems to be “replayed”, giving rise to some sort of presentness. However, the perceptual specious present is undoubtedly a standard case of specious present, and an analysis of it will also help to shed light on the nature of specious present in general.

5.2. *Preparing explananda: Unity, Continuity, and Immediacy*

In this section, I prepare phenomenological explananda of the specious present, which will then be explained in the next section by the predictive processing account. The list is: Unity, Presentness, Continuity, Saddleback-shape, Framelessness, and Immediacy. I will explain these items via the *Symphony No. 5* example. When you hear the first motif of "dit-dit-dit-dah", the tones are experienced in one specious present; the four tones are experienced *together*. In other words, they are *experientially united*. Thus, we get the following thesis.

> The Diachronic Unity Thesis:
> simultaneous contents can be experienced together, but so too can contents that appear to be successive (at least over short intervals). (Dainton, forthcoming)

The synchronic version of unity is common and more tractable (Bayne, 2010). For example, when I look at a computer screen while writing this chapter, I experience the computer screen and the keyboard together. Thus, my haptic experience of the keyboard is unified with my visual experience of the screen. This is a synchronic version of unity of consciousness. But what we have in the “dit-dit-dit-dah” example is a diachronic version; the experiences *within one specious present* are unified with each other despite short-intervals between them. The experiences within specious present do seem unified, but at the same time, experiences that are more distant in time do not seem unified. For example, experiences of dit, dit, dit are unified with dah, but the experience of the first dit is not unified with experience of the tones in the symphony 30 seconds later! We have to explain not only how experiences are diachronically united but also how experiences are *disunited*.

The next conspicuous characteristic of the specious present is *Presentness*. As I touched on in the introduction, conscious perception "flags now." The events in the world are experientially presented as happening in the *present* in conscious perception. It is noteworthy that the presentness in question is presentness in the *content* of perception. There is a sense in which you are aware of the fact you are *perceiving* now, but this introspective awareness should be distinguished from presentness in the content of perception. The presentness here relates to the openness to the world mentioned at the start of this chapter, and is thus more to do with our direct awareness of the world.

The next important characteristic is *Continuity*. As stated in the Diachronic Unity Thesis, the contents of perception simultaneously appear to be unified and successive. There is a temporal order or flow in experience—the first dit is followed by another dit, which is followed by the last dit, before ending in a dah. The content in the specious present also has a temporal order and it forms part of the larger flow or stream of conscious.

Order in the specious present is asymmetrical; the relation between two tones is different from a normal synchronic unity relation. Conversely, if your haptic experience of a keyboard is united with visual experience of a screen, your visual experience also is united with your haptic experience!

There are two aspects of Continuity here that should be distinguished: the succession of experiences and experience of succession. The succession of experiences refers to the *ontological* fact of continuity of experiences. For example, throughout a whole short-interval episode of drinking tea from a cup on the table, there is a chain of experiences—the visual experience of a cup, the tactile experience of holding a cup, the somatosensory experience of bringing it up to your lips, the gustatory experience of tea, and another somatosensory experience of bringing it down to the table. It is intuitive to think there is a chain (or stream) of conscious experiences without a gap. However, this cannot be established only from a subjective perspective. The fact that there is no experience of a gap falls short of proving that there is no gap in the experience; there might, as Dennett (1991) argues, be a blindness of with regard to these gaps. Thus this ontological continuity is not evident.

Conversely, however, it is undeniable that there is a *subjective* awareness of the continuity of experiences. During the tea-drinking episode, you would probably feel successiveness experiences. It seems like your visual experience of the cup "flows into" your tactile experience of holding the cup, for example. With regard to specious present as discussed in this chapter, there is a subjective successiveness within the specious present and a subjective successiveness that smears backwards and forwards from the specious present. Subjective successiveness within the specious present is what you feel when you hear "dit-dit-dit-dah." But, you would also feel a succession with past and future tones. This is subjective successiveness smearing from the specious present (this point is also related to *Framelessness*). Moreover, there are cases where you feel subjective successiveness less. For example, if you suddenly hear a loud, sudden "bang!" while listening to *Symphony No. 5*, the "bang!" sound would feel less continuous with your past experiences (and also probably with your future experiences). The subjective successiveness at stake here is mainly a smearing subjective successiveness—as it is likely that there is still a subjective successiveness within the experience of the "bang!" sound. This kind of "solitary" specious present also seems possible. Thus the presence or absence of subjective successiveness also wants explanation.

Ontological and subjective continuity do not always hang together. For example, when I park my car and walk to my apartment, there is a whole chain of experiences: turning off the engine, opening the car door, closing the car door, walking to the entrance, unlocking the front door etc. Let us also grant there is no gap in the chain of experiences. Even though the experiences are ontologically continuous, my experience of turning off the engine and my experience of opening the entrance door don't feel successive; they are too far apart for this. Thus, only some successive experiences are experienced as successive. Hence, as the famous Jamesian slogan goes, "a succession of feelings, in and of itself, is not a feeling of succession." Given this distinction, what *Continuity* as a phenomenological thesis has to explain is subjective continuity.

However, what exactly is an experience of succession that is not captured by unity or gap-free stream of consciousness? This most cardinal but most elusive phenomenal aspect of temporal consciousness needs independent treatment, and so I will further analyze this in the next section. Also, I reiterate that the experience of succession and the present aspect of the experience of the specious present just discussed is in tension, as noted in introduction—how can content experienced as present also be temporally successive? This puzzling aspect also has to be explicated.

The fourth characteristic of the specious present is *Saddleback-shape*. *Saddleback-shape* is a subtle yet important aspect of the specious present. There are two ways to think about the presentness of the contents of the specious present. One way is to think all the experiences within the specious present exhibit the same degree of presentness. As long as they are within the same specious present, the contents are experienced together, and so the contents are equally experienced as present; the first, second, and third dits are all equally experienced as happening now. This view also implies that the specious present has a rather sharp boundary, because beyond the specious present would be marked by the absence of this "flat" presentness. I call this the "plateau view", as all the contents experienced within the same specious present have the same "high" level of presentness and, while those experiences outside of specious present have none.

The other way is to think there are degrees of presentness within the specious present. In other words, there is a maximally present point (or at maximally present narrow band) within the specious present, with all other experiences being less "present". This view is well expressed by James, who I shall quote again:

> [T]he practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. (James, 1890/2013, p. 609)

The specious present has a saddleback shape—there is the highest point where "we can sit perched" and from there presentness gradually tails off in both directions. So, for example, when we experience "dit, dit, dit", all the dits are not experienced as equally present. For example, the second dit might be experienced as "maximally" present, with the first dit being experienced as a little bit in the past and the third dit being a little bit in the future.

I think this saddleback view is more phenomenologically appropriate. Consider another case: when we see a bird flying, the present position of the bird is specified, but we also experience its recent past position and the near future position. In addition to that, the plateau view suffers in accommodating subjective continuity (the experience of succession). How can the contents of experience be successive if all the contents in the specious present are equally experienced as "now"? Moreover, the plateau view suggests the present is isolated from past and future, thus it makes it difficult for how the subjective time flows.

There are two ways in which the boundary of the specious present could be, according to this saddleback view. Firstly, the saddleback view is consistent with the specious present having a sharp boundary; there might be a presentness cut-off line for an experience to be within specious present. If the content is about a time too far from now, it is not experienced as present. This suggestion seems to have some intuitive plausibility. However, the boundary need not be crisp. The content gradually might fall away from the specious present as the presentness of content gradually tails off. So the second possibility is that the boundary is fuzzy. I think the latter view is more phenomenologically apt because of Framelessness, which I will argue for next. The unavailability of this second option of fuzzy boundaries for the plateau view lends more support to the saddleback view.

The framelessness mentioned above is the fifth aspect of specious present to be explained. Our daily experiences seem like a stream—starting from the experience of waking up, the experience of washing teeth and so on—but each specious present that comprises the stream of consciousness does not seem to have a crisp boundary; it is not clear if a specific experience belongs to one specious present or other one. This point is well expressed in Metzinger,

> What is so hard to describe is the strong degree of integration holding between the experience of presence and the continuous conscious representation of change and duration. It is not as if you see the clouds drifting through a window, the window of the Now. There is no window frame. It is not as if the Now would be an island emerging in a river, in the continuous flow of consciously experienced events, as it were—in a strange way the island is a part of the river itself. (Metzinger 2004, 153, also quoted in Hohwy, Paton, & Palmer, 2015).

Again, auditory experience would be a most illustrative experience. When you hear music, we only hear limited tones in the specious present, yet we end up listening to a whole piece of music. However, it is not clear how many tones are experienced in "one" specious present, nor how one specious present should be demarcated from another. Without informing us of its boundary, the specious present flows. This characteristic obviously speaks to fuzzy boundaries and the saddleback-shape of specious present.

Last but not least, the sixth characteristic is Immediacy. Dainton formulates Immediacy this way:

> The Immediacy Thesis: change, succession and persistence can feature in our experience with the same vivid immediacy as colour or sound, or any other phenomenal feature. (Dainton, forthcoming)

The term “vivid immediacy” seems misleading, as the experience of change, succession, and persistence is not as vivid as direct sensory experiences such as color or sound. How we can understand change is a perennial philosophical question, as is the experience of change. You will likely stumble if asked what motion is apart from the collection of positions an object takes; motion per se is a subtle, elusive aspect of phenomenology. However, the point made here can be stated differently. Immediacy is the noninferential phenomenological character of the experience. In most cases, we do not feel like we are making inference about change or succession. When a batter sees a ball thrown from a pitcher at a baseball game, the batter wouldn’t feel as if he is making inference regarding whether the ball is moving toward him. If he tried to do this, his bat wouldn’t contact the ball. Rather, his experience of the ball moving towards him feels like direct perception, just like perception of red or perception of square shape. This does not exclude the possibility of unconscious inference, indeed, I will argue it is unconscious inference, but this immediate aspect of experience should be addressed.

Before moving on to the next section, here is the summary of the explananda that an adequate account of specious present has to address:

I. Unity: The contents of experience are experienced together.
II. Presentness: The contents of experience are about the present.
III. Continuity: The contents of the specious present feel successive.
IV. Saddleback-shape: Presentness within the specious present has a saddleback-shape: there is a peak in presentness with the degree of presentness tailing off both in the direction of the past and the future.
V. Framelessness: The boundary of the specious present is fuzzy: whether specific content belongs to one specious present or another is not always accessible to the subject.

VI. Immediacy: The experience of change, succession and persistence is as immediate as other uncontroversial kinds of experience such as color or sound.

5.3. *The experience of successions*

In the previous section, I argued that the gapless nature of our stream of consciousness is not enough to secure the experience of succession. However, what exactly is the experience of succession?

I will now elucidate the experience of succession by further analyzing the unity of consciousness, and the nested, hierarchical phenomenological structure of our stream of consciousness. By doing so, we can tie the notion of the diachronic unity of consciousness and *Continuity*. In the previous section, I argued that the contents of experience of the specious present are unified. Every part of the "dit-dit-dit-dah" of the opening of *Symphony No. 5* is experienced together. This diachronic unity is not analyzed by Dainton in further primitive terms, however, we can gain some useful ideas from Tim Bayne's analysis of *synchronic* unity of consciousness. Bayne argues synchronic unity can be conceived in terms of mereology.

> [T]wo conscious states are phenomenally unified when, and only when, they are co-subsumed. What it is to experience a headache and the sound of a trumpet together—what it is for these two experiences to possess a 'conjoint phenomenal character'—is for there to be a single experience that in some way includes both the experience of the headache and that of the trumpet. (Bayne, 2010, p. 20)

Can we give a similar analysis of diachronic unity? It seems we can. A straightforward answer is that the three experiences of the "dit"s and the experience of the "dah" are experienced together by virtue of being part of the experience of the opening motif of *Symphony No. 5*. Put more generally, two experiences are diachronically unified when, and only when they are subsumed by the experience of an event of a longer time-scale. Just as the visual experience of eyes is subsumed by an experience of the face, the experience of smaller regions more generally is subsumed by the experience of larger regions. In the *Symphony No. 5*., the motif is united with other motifs, and by virtue of that their experiences are also unified. In this way, experience has a nested, hierarchical structure. This aspect of experience is not often discussed, but Metzinger points to this aspect:

> If events are not only represented as being in temporal succession but are integrated into temporal figures (e.g., the extended gestalt of a consciously experienced musical motive), then a present emerges, because these events are now internally connected. They are not isolated atoms anymore, because they form a context for each other. Just as in visual perception different global stimulus properties—for instance, colors, shapes, and surface textures—are bound into a subjectively experienced object of perception (e.g., a consciously seen apple) in time perception as well, something like object formation takes place in which isolated events are integrated into a Now. (Metzinger, 2004, pp. 126-7)

> In addition, […] phenomenal wholes do not coexist as isolated entities, but appear as flexible, nested patterns or multilayered experiential gestalts. They form mereological hierarchies. Nestedness (or "convolution") is a property of any hierarchical system having entities of smaller scale enclosed within those of larger scale. (Metzinger, 2004, p. 143)

In one way, subsumption comes cheaply. When you are consciously seeing a white coffee mug in front of you while experiencing mild itchiness in your foot, you can make an aribitrary conjoint experience of the visual experience of the coffee mug and the tactile experience of the itchiness to subsume those two experiences. However, the content of a larger experience needs to concern a larger space and a longer timescale than those of the smaller experiences it subsumes.

In the case of a conjunction of the visual experience of a coffee mug and the tactile experience of itchiness, there is nothing more than that a conjoint experience involving a larger space and longer timescale. However, there are cases where a meaningful relationship is obtained between a larger experience and the smaller experiences subsumed. For example, the experience of a face might subsume the experience of eyes, as they have to occupy certain location in the face. Similarly, to be a part of experience of opening of *Symphony No. 5.*, the “dah” sound has to follow certain pattern in the sequences. It cannot come at the beginning or in between “dit”s.

In these examples, the larger experience mandates that the smaller experiences follow a certain pattern to subsumed. However, the larger subsuming experience can be something arbitrary. So, what is the difference between when there is meaningful relationship between the larger experience and the smaller experiences, and when there is none? I argue that when there is some meaningful relationship between the larger experience and the smaller experiences, the larger experience can have structured richer content. In Chapter 4 (“Poverty”), I argued that through a higher level model about the gist of a scene, we can experience richer details at a lower level. The same sort of mechanism operates here. The higher level model here corresponds to the larger, subsuming experience, and lower level model corresponds to the smaller experiences. I will explain this in more detail in Section 5.4.

We can see this in a diachronic case. (Almost) everyone knows how *Symphony No. 5.* starts, and so the opening “dit-dit-dit-dah” is experienced in one specious present. But when you first hear a musical piece, the development of music is less predictable; the parts of an unfamiliar musical piece feel disunified when we lack knowledge about the direction of the music. In other words, when we lack contextual knowledge, the specious present becomes shorter and the tones are experienced in a more solitary way. Then, as the music becomes more familiar it starts to flow smoothly. Thus the length of the specious present shrinks or stretches depending on how well structured the specious present is (more on this in the next section).

To summarize the minimal rule of subsumption would be stated as follows:

> Subsumption rule:
> For any two or more experiences to be subsumed by a larger experience, the larger experience has to involve a larger space and longer timescale than the smaller experiences.

Note that this subsumption relation between a larger experience and the smaller experiences is a synchronic one. The experiences need to coexist synchronically for there to be subsumption. Just as the experience as of a face has to be there at the same time as the experience of eyes in order to subsume the experience of the eyes, the larger experience of *Symphony No. 5*. has to be there to subsume the experience of "dit-dit-dit-dah". Because the larger and the smaller experiences have to exist concurrently, there are two possible versions of mereological analysis of the diachronic unity of consciousness. One version is a realistic version of diachronic unity; the other is a deflationary, reductive one.

According to the realistic version of diachronic unity, the three experiences of "dit"s and the experience of the "dah" are unified by virtue of being proper parts of the experience of the opening motif of *Symphony No. 5*. The formal definition of Realistic Diachronic Unity is as follows:

> Realistic Diachronic Unity:
> Two conscious states are diachronically unified when and only when they are both subsumed by the same larger experience.

However, this view is not palatable for those who think that past and future experiences do not exist and cannot possibly be unified with the current experiences. This position leads to a deflationary version of diachronic unity in which the only unified states are current states. The only unified states are synchronically existent states: the larger and the smaller experiences. Thus the experiences of "dit", "dit", "dit", and "dah" are not unified, as they cannot occur at the same time. Rather, the unified states are the experience of the opening motif of *Symphony No. 5.* and that of one "dit," for example. We cannot experience "dit", "dit", "dit", and "dah" together literally, but we can have an experience of each sound in and of itself, as well and an experience as of the opening motif. Thus, this view explains apparently "horizontal" diachronic unity by a simultaneous, "vertical" unity. Here, it is interesting to consider the *synchronic* analog of this deflationary view. When we experience a person's face, it is intuitive to think that we experience all the parts of the face and that these experiences are unified. We experience eyes, lips, a nose, and so on. However, it is now well known that we cannot simultaneously experience all these details, as only a few items attended are experienced (as was argued in Chapter 4). Therefore, we cannot possibly have unified experiences of all the details. One solution to this problem is to argue that we experience the object as a face while experiencing only a few parts of it as parts. As long as it is experienced as a face, this implies the object has "face-like" features and so accounts for the impression of experiencing all the details. In the same token, as long as the "dit-dit-dit-dah" it is experienced as the opening motif of *Symphony No, 5*, this implies that the "dit"s and the "dah" form a sequence. Put more generally, this Deflationary Diachronic Unity can be defined as:

> Deflationary Diachronic Unity:[41]
> When there are apparently diachronically unified conscious states, there are at least two synchronic state, one of which subsumes the other.

Deflationary Diachronic Unity fits better with the Saddleback-shape discussed in the previous section than Realistic Diachronic Unity. Realistic Diachronic Unity provides no reasons for a peak in presentness, whereas this deflationary version naturally explains its saddleback shape because, under this view, what actually exists is the current smaller experience along with the larger experience that subsumes it (the experience of one "dit" and the experience of the opening motif for example), and real successive experiences are not required. They are only implicated by the larger experience. Thus, I will hereafter support Deflationary Diachronic Unity, but I will also touch on the realist version when it is needed.

In this section, I explained subjective continuity by tying it to diachronic unity. From this, we get two versions of unity. In Section 5.5., I will argue for the deflationary account, however first, I will look at what the predictive processing account adds to our understanding of the specious present.

5.4. *A Predictive processing account of the specious present*

This section gives an account of the phenomenological explananda discussed in the previous section. The predictive processing account, by invoking predictive processing mechanisms in the brain, explains why we have the relevant temporal phenomenology for our experience of the present. To begin with, the analysis from the previous section clearly indicates a hierarchically organized system in our "present" experiences. Thus, one component of predictive processing marshaled for explanation is *Hierarchy*. We

[41] The most deflationary conception of diachronic unity would be one that denies the existence of diachronic unity itself. But, in this chapter, I will set aside this position and assume the realism of diachronic unity.

will also need *Prediction Error Minimization* and *Precision Expectation*. According to *Prediction Error Minimization*, the brain constantly engages in predicting hypotheses about the world from incoming sensory stimulation. If there is mismatch between predicted sensory stimulation and actual sensory stimulation, there is a prediction error. The brain amends its hypothesis, thereby minimizing error. The hypotheses are crucially about the worldly causes of sensory stimulation, therefore any given hypothesis is about existing things or events unfolding in the world "now". The "now" has some width: an event unfolds over time so the "now" includes both the recent past and the predicted future. This already explains Presentness: the contents of the prediction error minimization system are about the present.

Hierarchy, together with *Prediction Error Minimization*, explains *Unity* and *Continuity*. According to *Hierarchy*, to minimize prediction errors, the brain is organized hierarchically. Higher levels in the hierarchy are dedicated to longer timescales whereas levels lower in the hierarchy deal with shorter timescales. At level n+1, a prediction is made about what happened at level n, and prediction errors that cannot be resolved at level n are sent up to level $n+1$ for resolution at higher levels. This *Hierarchy* clearly has some echoes with the hierarchical analysis of the unity of consciousness. According to the considerations raised in the previous section, for both realistic and deflationary diachronic unity, the subsumption of smaller experiences by larger experiences needs follow the subsumption rule discussed in this chapter.

This is precisely what the hierarchical system of the predictive processing model predicts. In this model, the higher level models predict patterns at the lower levels, and this *Hierarchy* can explain both *Realistic Diachronic Unity* and *Deflationary Diachronic Unity.*

The higher level model corresponds to a broader experience and the lower level model corresponds to a narrower experience. How much the overall experience can be structured and nested is determined by how well lower levels are predicted. For example, the higher level of the opening motif of the symphony predicts prediction errors at the lower level, when there is a sequence of "dit", "dit", "dit", and "dah" at that level. If *Realistic Diachronic Unity* is right, all the smaller experiences of "dit", "dit", "dit", and "dah" and the larger experience of the opening motif of the symphony are experienced. Conversely, if *Deflationary Diachronic Unity* is right, only one of the smaller experiences among the experiences of "dit", "dit", "dit", and "dah" is experienced at one time, but this is simultaneous with the broader experience of the motif being experienced.

Saddleback shape, Tension between presentness and the experience of succession and the question of *how successive can contents be represented by a momentary experience* can also be answered also by *Hierarchy and Prediction Error Minimisation*. *Hierarchy* in the brain indicates that at a single moment, experiences of different timescales are experienced synchronically. One can experience *Symphony No. 5*., the opening motif of *Symphony No. 5*., and the "dah" at the same time. They are all about the present, but they concern different timescales.

To entertain the model of the opening motif of *Symphony No. 5*. without prediction errors is to have lower-level models of "dit" "dit" "dit" "dah" in this order. The higher-levels concern a longer now and the lower-levels concern a shorter now, and they are superimposed. Contents change rapidly at lower-level, for instance, when "dah" is represented at a lower-level, none of three "dit"s are represented. But by virtue of being subsumed by a larger experience (that is predicted by longer timescale model), the experience of succession is created. *Realistic Diachronic Unity* requires actual sequences of lower-level models but *Deflationary Diachronic Unity* doesn't. If *Deflationary Diachronic Unity* is true, the three "dit"s are *retrodictions* created by the higher-level model of the opening motif of *Symphony No. 5*. That is, we never, strictly speaking, hear "dit-dit-dit-dah" as happening now.

Thus, the puzzles concerning presentness and successiveness are solved by introducing representations of "now" of different lengths:

> Longer now: I am now hearing the opening motif of *Symphony No. 5*.
> Shorter now: I am now hearing "dah", which follows a "dit."

What is represented at any given time is only models concerning different timescales. Thus I can have the model that I am now hearing the opening motif of Symphony No. 5. and the model that I am now hearing "dah", this follows a "dit", but the interplay between higher levels and lower levels creates a succession at the lower level (Figure 10). You might question the difference between hearing opening motif of *Symphony No. 5*. and hearing "dit-dit-dit-dah". If they are the same thing, we end up conceding that we can hear "dit-dit-dit-dah" as happening now. However, there is a difference between the two. Higher-level models concern longer timescales but tend to lack details, whereas lower-level models have higher levels of detail. In the case of the opening motif of *Symphony No. 5*, there are many ways the three “dit”s and the “dah” can be played, but the higher-level model of the opening motif of *Symphony No. 5* does not represent these details, all that matters is that it is the opening motif of *Symphony No. 5*. Conversely, the lower level can represent the detailed way of how the “dah” is played.

The situation here is similar to the one we discussed in the previous chapter. There I argued that poverty is overcome through the higher level model, but that this means that changes congruent with the higher level model tends to be overlooked. From this, we can predict a temporal version of change blindness. It should be possible that a subject is tricked because of higher level models: because of the higher expectation for higher level models, we might disregard prediction errors in lower level models.

There is and auditory version of change blindness. For example, in one study, a subject is required to shadow words read over headphones (Bayne, 2010). In the middle of the task, the identity of the speaker changes, but more than 40% of participants fail to detect the change. One interpretation of this study is that a subject fails to notice the change in detail of tone or pitch of the voice because of the higher expectation for continuity of the subject and lower precision expectation. This is not change blindness by and in itself, but the phenomenon discovered in this study can be explained by the same mechanism used to explain the specious present.

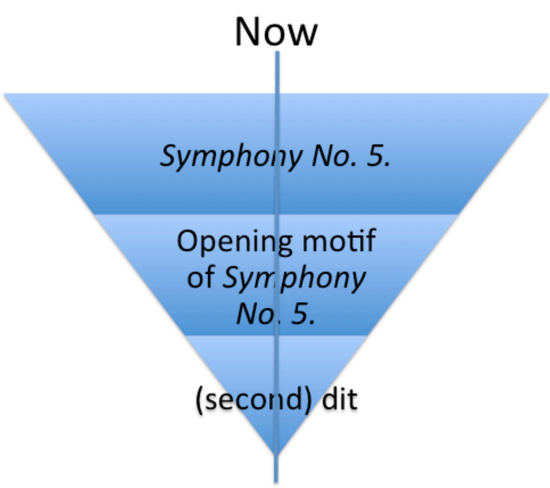


Figure 10 multiple now hypothesis

How can **Saddleback-shape** and **Framelessness** be explained in this setting? The inverse triangle figure in Figure 10 represents an upside-down **Saddleback-shape**. The inverse triangle reveals a shorter, narrower present, with higher layers in the hierarchy concerning a broader recent past and near future centered on that narrower present. The framelessness of the specious present—we don't see the border of specious present—is related to **Precision Expectation**. **Precision Expectation** is a second order statistical inference about the precision of prediction error. The representation as of a shorter "now"—in this case the experience of the (second) dit has a relatively higher precision because it is mainly driven by sensory stimulation. However, the preceding (first) "dit" (in Husserlian term, *retention*) is represented mainly through priors provided by the higher level models. The higher level models concern longer timescales, but they lack detail because they have lower precision. Because of this, the precision of priors at the lower level provided by the higher level models is estimated to be low. Moreover, since the precision of prediction error at the lower level is estimated to be low (simply because the sensory stimulation relevant to past tone is gone), the posterior distribution mostly relies on a prior that does not have high precision. The same story can be told about the following (third) "dit" (in Husserlian term, *protention*). The upshot of this is that the more distant (both in the past and in the future) the time of the content of the model becomes, the less precise the posterior distribution becomes. As a result, only the tones close to the current, narrower present are experienced. Thus, even though it looks seems, prima facie, as if the upper side of the triangle can be infinitely wide (for example, it might look as though we can even experience the change of the sunspots, which happens in the order of decades), this is not the case. Moreover, the way the estimated precision drops is not a discrete phenomenon because of the probabilistic nature of statistical inference. Thus we can predict gradual fade out of the contents from the specious present.

Before concluding this section, let's turn, finally, to Immediacy.

> *Immediacy***:** the experience of change, succession and persistence is as immediate as other uncontroversial kinds of experience such as color or sound.

I explained the experience of change, succession, and persistence in terms of hierarchical structure in experience. To experience succession is to experience events as parts of experiences of longer timer scale events. For example, experiencing the succession of "Dit-dit-dit-dah" is to be experiencing them as parts of opening motif of *Symphony No. 5*. The unity invoked here is not called upon in the explanation of color or sound, but the tools used to explain it are no different to the tools used to explain color and sound. The experience of change, succession, and persistence involves more nuanced interaction between levels but they are all explained by the predictive processing framework. The experiences exhibit immediacy because they are also the product of *unconscious* perceptual inference, just as with the experience of color or sound. Because the computational processes are unconscious, the resulting perception has immediacy.

5.5. *Does experience itself extend over time?*

In the previous section we saw how predictive processing provides the cognitive machinery that explains the phenomenological features of the specious present discussed in Section 2 and Section 3. In this section and the next, I will discuss issues regarding the timing of experience itself. In the beginning of this chapter, I mentioned the distinction between the time of events in the world, the time of apparent events in experience, and the time of experience itself. There is a debate about whether the experience of time-extended events in the specious present is itself time-extended. It is certainly intuitive to think that when we experience some event for a certain period of time, then the length of the experience of the event mirrors the event's duration. However, there are reasons to think the contrary. The experience itself is momentary but it can only represent time-extended events. The former view is called extentionalism, which is supported by Dainton (2006) whereas the latter is called retentionalism, supported by Husserl (1991) and Grush (2007b).

The definition of extentionalism is as follows:

> *Extensionalism*: Experiences that have time-extended contents are themselves extended over time.

Likewise, the definition of retentionalism can be given like this:

Retentionalism: Experiences that carry time-extended contents are themselves momentary but can represent time-extended contents.

In Section 3, we saw two versions of diachronic unity: realistic diachronic unity and deflationary diachronic unity. In a specious present, we seem to be able to experience extended events—we hear the "dit-dit-dit-dah," for example. On Realistic Diachronic Unity, we can have conscious mental representations of different times that make up a specious present. On Deflationary Diachronic Unity, we can only have synchronic unified conscious mental states. This roughly corresponds to the two views above about whether the experience of the specious present itself has extension.

In this section, I evaluate the arguments against extensionalism. My assessment is that the extensionalists can respond to almost all of the counter-arguments given here, but the cannot explain the *Saddleback-shape* discussed in the previous two sections. This is not itself a strong objection, but retentionalism can easily accommodate this aspect, which is a point in favor of retentionalism. We saw that predictive processing is particularly well suited to explaining *Saddleback-shape*—thus, it can be said that the predictive processing account supports retentionalism over extensionalism.

First, even though I claimed that it is intuitive to think that if we can experience temporally extended events, experience itself is time-extended, our intuition seems to also pull in the other direction. For example, when you see an apple dropping from a tree, the apple is now travelling through the air; it is no longer hanging from the branch. That the apple hung from a particular branch is in the past; it is not the case now. Since perception is perception of now, even though you experience an apple as dropping from a tree, that the apple hung from the tree previously cannot be perceived, it has to be in memory. In a nutshell, because the past does not exist, past experiences cannot take part in the present experience of succession. Thus, some people are more naturally inclined to subscribe to this principle:

Principle of Simultaneous Awareness (PSA): if one experiences succession or temporal structure at all, then one experiences it at a moment (Phillips, 2010, section 1)

This principle is shared by retentionalism and Deflationary Diachronic Unity. Thus intuitions about temporal extension pull in different directions, and so appeal to intuition cannot settle the issue. Apart from these intuitions, more sophisticated arguments can be provided. A classic example is in *Analysis of Mind* (1921), where Bertrand Russell presented the famous five-minutes hypothesis:

> [T]here is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that "remembered" a wholly unreal past. (Russell, 1921/2005, p. 159)

According to this hypothesis, we cannot distinguish, on the basis of memory and experience, between two hypotheses: the hypothesis that the world has been as it is thought to have been, and that the world was created five minutes ago with the illusion of having long existed. The five-minutes hypothesis itself is about the objective history of the world, but can be used as an argument for retentionalism. If the scenario that the world was created five minutes ago is logically possible, why cannot this be 0.1 seconds ago? Even if it were created 0.1 seconds ago, it seems that we cannot know this from experience and memory. It seems that be it five minutes ago or 0.1 seconds ago, a time-extended specious present does not require an actual past, let alone a past experience lasting two seconds, but only needs memory.

Perhaps, we don't even need to conceive of the creation of the whole world to establish this point. We can think of a version of the Swampman case (Davidson, 1987). In this thought experiment, Davidson is hit by lightning and the lightning completely dissolves him into his constituent molecules. However, by sheer chance, the lightning also rearranges nearby molecules in the exactly the same way that Davidson used to be arranged. Davidson names the creature Swampman and questions if Swampman possesses any intentional states. Davidson's answer is negative: Swampman does not have intentional states because it lacks the causal history that is required to have the appropriate kinds of intentional states.

His answer is controversial, but the relevant question for us here is whether Swampman can have a time-extended specious present. If Davidson was experiencing a leaf falling on the swamp, does Swampman also experience the time-extended event of the motion of a leaf in a specious present? It seems he does. If asked, Swampman would readily report that he sees a falling leaf, and he could act on the leaf if needed. Besides that, even though it is well accepted that the contents of intentional states are at least partly individuated by their causal history with the environment, it is far more common to think that consciousness (including temporal consciousness) only supervenes on the brain state (for example, Clark and Chalmers, 1998).

The arguments presented so far mainly rest on logical possibility or conceivability, which leaves room for disagreement. However, there is a more empirical argument: Lee's trace integration argument (2014). Lee looks at the actual mechanisms that an organism employs to detect motion. He mainly utilizes a simplified model, but the mechanisms are based on actual motion detection mechanism in a fly. In the case he uses, the detector is composed of four parts. Two detectors that are sensitive to the presence of an object at two nearby positions, an AND gate, and a delay filter that delays the transmission of signals from a detector to the AND gate. The AND gate fires only when it received signals from both detectors. The parts are arranged so that the AND gate fires mostly when there is a motion between certain locations at a certain speed. The two detectors are activated with a certain time lag, but because of the delay filter, the activations arrive simultaneously at the AND gate and because it received two signals, it fires. In this way, the firing of the AND gate can indicate certain motion. Lee's crucial claim is that subjective experience is realized at the stage after the firing of AND gate and that the successive events of activation of a sensory detector is not sufficient for a motion to be experienced. This comes from the relatively uncontroversial assumption that the content of conscious experience is available for executive control, such as verbal reports or rational control of action. For the information about motion to be available, the mere sensory stimulation at the level of detectors is clearly insufficient. The first stimulation at a detector is already lost by the time the other detector is activated by the moving object. Thus, the information about the location of an object has to be preserved in the brain. In other words, "the initial extended stimulation has to leave a trace in the brain" (Lee, 2014, p. 14). What is needed to experience motion is the integration of information; all the relevant pieces of information about time have to be available to the executive system *simultaneously*, in this case, information about the presence of an object at a location and the presence of the object at a nearby location, with a short time lag between the two. Therefore, Lee concludes that at the stage that the integration occurs, the experience of motion takes place. This means that even though events experienced have duration, experience itself does not have the same duration.

His argument does not appeal to conceivability of any sort. Rather, it hinges on an empirical observation about the kind of mechanism needed for organisms to represent motion in this world. Lee says,

> Models of temporal computation implicitly assume that setting up such traces and then simultaneously integrating them is the task that the brain has to perform. (Lee, 2014, p. 15)

I think this observation is an important one, but there is a pitfall. The claim is plausible to the extent that the combination in a moment of information about the recent past registered in the brain and the newer information received from sensory organs is necessary for an organism to experience a time-extended specious present. However, this might not be sufficient. The response given by extensionalist Barry Dainton to the objection from temporal illusion precisely rejects this sufficiency. The objection from temporal illusion to extensionalism is a well-organized one, so I will look into it in the next section.

5.6. *Temporal illusion and consciousness*

The temporal illusion argument against extensionalism claims that extensionalism cannot accommodate important kinds of temporal illusion I argue that this objection rests on the false assumption that extensionalists have to embrace the naïve realistic claim that there is a close match between temporal properties of the world events and those of experience.

The relevant illusions are the cutaneous rabbit illusion and apparent motion. In cutaneous rabbit illusion, two sets of successive two-taps are delivered at different positions on the forearm in a short time interval (Geldard & Sherrick, 1972). One set of two-taps is usually given at near wrist and the other is given at near elbow. This, if done correctly, generates the illusion of four successive taps gradually moving up from wrist to elbow—it feels as if a "rabbit" is crawling up your forearm. The interesting temporal aspect of this phenomenon is related to the location of the second tap. The felt location of second tap is influenced by the taps at near elbow, but when the second tap is delivered, the taps near elbow have not yet been delivered! This suggests that the brain does not produce a conscious percept as the sensory stimulation arrives. Rather, there is a “temporal window” in which sensory stimulation is processed and interpreted together to make a conscious percept. Therefore, there is a processing delay because of the temporal window.

Another similar kind of temporal illusion is apparent motion. Apparent motion denotes phenomena where the successive presentation of static images produces the perception of motion. For example, in color phi phenomena (Dennett (1991, chapter 6) argues for this phenomenon at length),[42] two dots of different colors located adjacently are successively flashed. Suppose a red dot located at P flashes first and a blue dot located at Q flashes soon after that. If things work successfully, motion of the dot from P to Q is perceived, and it changes its color from red to blue in between P and Q. Similarly to the cutaneous rabbit illusion, there is an intriguing temporal aspect. The perception of motion and of the color change at the middle point is produced because of the second flash of the blue dot. But when the color has changed in experience, the second flash has not yet been experienced! This also suggests the existence of a temporal window and a delay. The brain receives two flashes before committing to a percept, and judges that there was a movement of a dot changing color in the middle rather than that two dots of different colors located adjacently flashed one after another.

Grush (2007b) argues that temporal illusions of this kind support retentionalism rather than extensionalism. To understand why he thinks so, it is easier if we start from his own account: the trajectory estimation model. The core idea is that the time *of representation* does not have to closely match the time *of represented events*. The trajectory estimator is a representation about a temporally extended environmental trajectory, that is, even though the trajectory estimator itself is momentary, it is about temporally extended, successive events. Grush explains the cutaneous rabbit illusion as follows. Suppose at t=1, 2, and 3, taps are delivered, one at a time, near the wrist, and at t=4 and 5, taps are delivered near the elbow, also one at a time. How does a trajectory estimator represent this sequence of events? At t=3, the trajectory estimator estimates what the events in the world are like from t=-1 to t=7 (4 time-points in the direction of the past and future). At that time, only three taps at near wrist have been sensorily

[42] http://www.yorku.ca/eye/colorphi.htm

registered, thus the estimator only represents three taps there (again, one tap at a time). Also, because of the regularity of the taps, the brain predicts another tap at near wrist at t=4. At t=5, however, when two taps are delivered near the elbow, the content of the trajectory estimator changes. Once the two taps near the elbow are registered in addition to the three taps near the wrist, the brain judges that such a big movement from wrist to elbow is unlikely, and so the brain predicts a continual progressive movement from near the wrist to near the elbow from t=1 to t=5. Our phenomonelogy is explained by the content of trajectory estimator at t=5. Even though the representation itself is produced at t=5, it retrodicts events happened before the time of its production. Grush's account has many echoes with my predictive processing account developed later in this chapter, especially in that both Grush's account and my predictive processing account rely on a predictive mechanism. However, there are a few phenomenological and functional points that Grush's account does not cover. I will come back to these points later.

Why does Grush think the extensionalist cannot explain temporal illusions adequately? He considers two versions of extensionalism (Grush, 2007b, p. 14) which might explain apparent motion. On the first version, the content of an extended perceptual event just is the content of events in the world happening at the same period. Then, the content can be nothing but that a red dot flashes and then a blue dot flashes at a nearby location. On the other version, different chunks of an extended perceptual event embody different perceptual interpretations. A perceptual chunk only includes the first flash and the following blank. Then the perceptual interpretation for the chunk is that there is only a flash. Another perceptual chunk includes both the first flash and the second flash. The perceptual interpretation for this chunk would be that there is a dot moving from one location to another. Grush claims that this version is untenable because "at any given time there are many, possibly inconsistent, perceptual state." (Grush, 2007b, p. 15). For example the time between the first flash and the second flash is included in both of the perceptual chunk whose perceptual interpretation is that there is only a flash and the perceptual chunk whose perceptual interpretation is that there is a moving dot. These two perceptual interpretations are inconsistent. Ascribing two inconsistent perceptual contents at the same time is unpalatable and also betrays phenomenology; it does not seem to us that a dot is stationary and in motion at the same time. It is true that even on Grush's trajectory estimation model, the two contents of the trajectory estimator are inconsistent. But the trajectory estimator does not represent inconsistent contents *at the same time*. A change in the contents of beliefs or judgments is commonplace so it is not problematic. The root of this unwanted result seems to stem from the assumption that the time of experience closely matches with the time of events. If this is mandatory, perceptual representation about the time between the first flash and the second flash has to itself be in place between first flash and second flash. Thus even though the brain changes the interpretation of the event, the representation about that period also has to be in that period; it has to assign two representations of inconsistent contents at the same time. The trajectory estimator model, or any retentional model, doesn't subscribe to this constraint. The inconsistent contents can be held at different times, and thus the problem doesn't occur.

However, *contra* Grush, this matching relation between the time of experience and the time of events is not forced on the extensionalist. The definition of extentionalism was:

> **Extensionalism:** Experiences that carry time-extended contents are themselves extended over time.

There is nothing in this definition that prevents extensionalism from permitting a separation between the time of an experience and the time of the event-in-the-world. Extentionalism need subscribe to a naïve realism according to which the worldly time just is the time of experience. Extensionalism is rather a straightforward solution to the problem of the specious present: phenomenally, present perception can represent extended events, thus experience itself has extension. In other words, it does posit a mirroring relation from *represented* events to experience, but it does not posit a mirroring relation from *worldly* events to experience. Therefore, extensionalism can tolerate the gap between the time of experience and the time of worldly events. The response to this criticism given by Dainton (2006) also is along this line. He argues that there is a delay in processing between perceived events and the experience of the events, so the first flash is never consciously experienced alone. A conscious percept can thus be formed based on the first flash and the second flash, so the resulting percept is of a dot moving. Dainton argues that the experience of a moving dot is extended not momentary, but there is a delay from the actual event in the world. This response demonstrates that extensinalism is actually compatible with an interpretative mechanism in the brain.

This seems a promising response, yet there is another problem. This response is at odds with an apparently plausible principle about the relationship between experience and the realizer of the experience. In a related discussion, Lee (2014) argues for what he calls the Temporal Correlation Principle, according to which experiences that have the same realizer timing have the same timing of experience. The response given by Dainton seems to violate this principle. According to him, the representation of the red dot flashing and the blue dot flashing becomes conscious at the same time. They should therefore be experienced at the same time, and so it seems impossible for the experiences to take place successively.

How can an extensionalist give a rejoinder to this? A key observation is that an experience of a red dot and of a blue dot is *embedded* in the larger experience of movement of a dot that changes its color midway. It is different from experiencing a red dot on my right and experiencing a blue dot on my left at the same time. Phenomenology exhibits this kind of hierarchical structure or nested-ness: the larger experience (a dot is moving and it changes the color in the midway) gives context to the smaller experiences (a red dot flashing and a blue dot flashing), and influences the order of the smaller experiences. Therefore, even though it seems true that if there are no specific reasons to think one experience comes later than another, experiences that have the same realizer timing have the same timing, but if there is a context that the experiences take place in order, the principle does not seem to hold.

To summarize the debate surrounding temporal illusion, the argument against extensionalism fails. The objection rests on a false assumption that the timing of events closely matches with the timing of the experience. But it is not necessarily so. Extensionalists can allow for a delay.

Where does this debate take us? One advantage of retentionalism is its compatibility with Russell's five minute's hypothesis and the Swampman case. Another advantage the retentionalist has is a phenomenological one: compatibility with the *Saddleback-shape*. According to *Saddleback-shape*, there presentness comes in degrees. The retentionalist can explain the maximally present point by the momentary nature of experience. The retentionalist version of the predictive processing account also nicely accommodated this aspect. How can the extensionalist explain this aspect? It seems that extensionalism is ill-equipped to explain the difference. Thus until that aspect is explained by an extensionalist, the retentionalism is in a superior position.

5.7. *Answering four problems*

Before moving onto the conclusion, let us summarize the answers to the four problems raised in Chapter 1.

Answering the inverse problem

The inverse problem for this chapter is how we can experience the specious present given that sensory stimulation at a time is compatible with many different past and future contents. I argued that retentional content (recent past) and protentional content (recent future) relies on priors provided by higher-level models.

Answering the description problem

I offered an analysis of the phenomenology of the specious present in Section 5.2. Many phenomenological features discovered by introspection are explained by the predictive processing account, including *Saddleback-shape, Immediacy, and Framelessness*. They are all subtle aspects of phenomenology—predictive processing gives strength to phenomenological analysis by providing cognitive mechanisms that underlie such phenomenology.

Answering the perceptuality problem

The perceptuality of the specious present can be dealt with in a similar way to Poverty. I argued that the experience of the specious present involves the overlap of multiple "now"s differing in the timescale they deal with. This is akin to arguing that the content of perception (at a time) is the conjoint of gist perception and more detailed perception (and prediction from gist). In both cases, models at different levels jointly explain away sensory inputs—thus they are all perceptual according to our criterion for perceptuality.

Answering the reason problem

The hierarchical Bayesian inference also neatly explains why we have the relevant phenomenology. As we have seen, *Hierarchy* played a central role in explaining why we experience a specious present which has duration. Moreover, *Precision Expectation* also explained why we cannot experience an infinitely long specious present: what we are receiving is sensory stimulation at that time point, and contents pertaining to the too distant past and future are not expected to be predictable with high precision.

5.8. *Conclusion*

In this chapter, the target of our inquiry was the specious present. Our perceptual experience is as of now—and at the same time the experienced "now" has certain "breadth." We have seen how the predictive processing account explains many phenomenological features of the specious present, including the puzzling experience of succession. The predictive processing account also predicts change blindness within a specious present. The exact phenomenon has not been found yet, but a similar phenomenon has been found: auditory change blindness, and this would likely be a promising area of research.

We also discussed the metaphysical issue of the temporal property of experience itself as opposed to the temporal property of the content of experience. Even though a conclusive argument was not given, it was found that phenomenological observation of the *Saddleback-shape* phenomenon supports retentionalism. The Saddleback-shape was well accounted for by the predictive processing view, and thus the predictive processing version of rententionalism emerges as the best account for both the phenomenology and the metaphysics of the specious present.

6. *The fourth P: Particularity*

6.1. *Introduction*

The fourth P is Particularity. Let us again start with an example. When you are having an ordinary perceptual experience, such as visual experience of a mug of beer at a bar, your perceptual experience is as of the *particular* beer mug that exists in front of you—it is not as of a *generic* beer that does not occupy spatiotemporal location in the world. This contrasts with, for example, the number nine, which does not occupy a spatiotemporal location. This fact also seems to be reflected in your phenomenology: it *seems* to you that the experience is of a particular object. In having an experience of a beer mug, it phenomenologically seems to you that you are seeing a particular beer mug that exists in your vicinity. The focus of this chapter is that aspect of perceptual phenomenology: phenomenal particularity. However, first I will need to argue particularity is exhibited in phenomenology. To see how this is debatable, let us consider the following scenario:

> Swapping scenario
>
> There is a red tomato on the table and you are looking at it. You are having a visual experience of a red spherical object. But the tomato is now replaced, unbeknownst to you, with a different tomato that is qualitatively indistinguishable. You did not notice the swap taking place, and your visual experience seems unchanged—at least to you, the subject of the experience.

This scenario suggests that perceptual phenomenology does not concern the particular object the perceptual experience is about; experience might concern what properties objects have, but it does not concern the numerical identity of objects itself. Upon facing this scenario, you might make a concession and argue that particularity is not reflected in phenomenology but it is reflected in other aspects of perceptual experience. For example, one can argue that the particularity of objects is reflected in the content of experience, even though the difference is not noticeable to the subject. This is an approach taken by Gareth Evans (1982) and John McDowell (1986), but doing so amounts to overlooking phenomenal particularity. Moreover, their strategy of incorporating particularity into content turns the notion of content into an externalist one. Even though the subject has no knowledge of the replacement of the object in the swapping scenario, the view of people like Evans and McDowell ascribes different contents to pre-swap experience and post-swap experience because there are different particulars. This conception of content is possible, but the targets of my thesis are phenomenological ones. I will therefore not discuss the validity of this externalist conception of content. What I focus on here is instead the phenomenological aspect of particularity and a corresponding conception of the content, which is compatible with the swapping scenario. Endorsing accounts given by Burge (2010) and Montague (2011) I argue this phenomenological particularity can be accounted for by demonstrative elements in content.

This chapter has two parts. In the first part, I provide a philosophical analysis of phenomenal particularity. I start with an examination of different philosophical accounts of particularity and I concur with the ones given by Burge (2010) and Montague (2011). As mentioned above, these philosophers think perceptual content involves the demonstrative element. Assuming this view, I then go on to argue that there are two kinds of phenomenal particularity: narrow and broad particularity, which are sometimes confused in the debate. I also argue there is an aspect of perceptual experience that does not exhibit particularity at all, namely, gist perception. This first half of chapter is also an answer to the description problem introduced in Chapter 1. Unlike my strategy in other chapters, the answer I propose to the description problem in this chapter is arrived at almost exclusively through philosophical analysis; I do not make much use of empirical evidence or theories. As we will see later in the chapter, this conception of phenomenological particularity presents an obstacle to predictive processing.

In the second part, I explain phenomenal particularity as conceived from the viewpoint of predictive processing. As I set out in Chapter 1, predictive processing is expected to have the resources to explain phenomenal particularity. At first glance, phenomenal particularity seems easy to accommodate from inferential views of perception such as predictive processing approaches. However, on closer inspection, a problem is easily recognized: at least some perceptual representations involving phenomenal particularity are not responsive to beliefs. In this chapter, I try to solve this problem and give an account of phenomenal particularity from the viewpoint of predictive processing. Finally, I will discuss the three other problems I raised in Chapter 1: the perceptuality problem, the inverse problem, and the reason problem. To begin with, in the next section, I set the stage for the chapter.

6.2. *Phenomenal content and particularity*

As I mentioned in the introduction, phenomenal particularity is reflected in perceptual content. It is related to one of the prominent characteristics of perception: openness to the world (McDowell, 1994, p. 111). Perceptual experience directly presents us with mind-independent objects. The objects are also presented as *particular* objects that are present *here* and *now*. But this openness is just a seeming, and can be misleading. For example, even when it seems to you that there is a tomato, one might not actually be there. In such a case, the experience is non-verdical. Your experience would also correctly represent the world sometimes. When it seems to you that there is a tomato and there actually is a tomato, the visual experience is veridical. I call this aspect of perceptual experience which is assessable in terms of veridicality *phenomenal content*, or simply *content* when there is no danger of confusion. Under this conception of content, content is solely internalist; it is something "the subject has access to." Thus, if two experiences are indistinguishable to the subject, they have the same content.
This conception of content is explicitly embraced by Martin Davies:

> [I]n the case of perceptual content, it is plausible that if two objects are genuinely indistinguishable for a subject, then a perceptual experience of one has the same content as a perceptual experience of the other. The source of this plausibility is the thought that the perceptual content of experience is a phenomenal notion: perceptual content is a matter of how the world seems to the experiencer [...]. *If perceptual content is, in this sense, 'phenomenological content'* [emphasis added] [...] then, where there is no phenomenological difference for the subject, there is no difference in perceptual content. (Davies, 1992, p. 26, *emphasis added*)

He proceeds to argue that the "format" of this kind of content is not object-involving, given that two different particular objects can give rise to same phenomenology:

> If perceptual content is phenomenological content, then, it seems, it is not object-involving. But from this it does not follow that perceptual content is

> not truth conditional—not fully representational; for we can take perceptual content to be *existentially quantified content* [emphasis added]. A visual experience may present the world as containing an object of a certain size and shape, in a certain direction, at a certain distance from the subject. (Davies, 1992, p. 26)

In a similar vein, Colin McGinn argues:

> It follows from what we have just established that the content of experience is not to be specified by using terms that refer to the object of experience, on pain of denying that distinct objects can seem precisely the same: so when we are describing the content of an experience we should not make singular reference to the object of the experience in the context following 'as of'. In fact it seems right to uphold a stronger thesis about experiential content; that an accurate description of the phenomenological content of an experience will employ *only general terms* [emphasis added] to specify how the experience represents the world. (McGinn, 1982, p. 51)

Put together, these positions mean that we can refer to objects only through descriptions that are comprised of terms about general attributes and existential quantifiers. If content is characterized only by general terms such as red or spherical, two experiences pre- and post-swapping would have exactly the same content. Both would be characterized as "there is a red spherical object in Location *L*".

However, this move is not mandatory. It is correct that if contents are characterized using only general terms and existential quantifiers, a subject would not notice the difference in identity in the swapping scenario, but this amounts to overlooking phenomenal particularity. Not only do we perceive a particular object in the surrounding environment (in successful cases), it also phenomenologically seems to us that we perceive *this* object rather than other random objects. This suggests that singular terms (more specifically, demonstratives) are involved in phenomenal content. Thus, a position that accepts phenomenal particularity would be a position that claims that singular terms are used to characterize phenomenal content. In the next section, I will examine some existing attempts to account for particularity. Some of these attempts are successful in addressing phenomenal particularity, others are not.

6.3. *Attempts to explain particularity*

In this section, I evaluate existing attempts to explain particularity. I will start with an attempt that is not compatible with phenomenal particularity. One straightforward way to account for particularity is to incorporate particular objects of perception as *constituents* of perceptual content. This approach claims the actual apple in the world constitutes the content of your conscious perception. I will discuss this position first and argue that is approach is not an apt account of phenomenal content. I will then argue for another kind of approach, naïve realist approach, which involves singular terms and takes the phenomenal aspect of particularity seriously.

Gareth Evans and John McDowell are among the philosophers who take the first approach. They think that in successful cases of perception, content comprises actual objects. This entails that even in hallucinatory cases that the subject cannot distinguish from perceptions, the content is fundamentally different from those of perceptions. Indeed, talking about hallucinatory cases, McDowell (1986, p. 165), states, "these 'contents' could not yield answers to the question what it is that someone thinks; there is really no reason to recognize them as contents at all". Evans (1982, p. 46) makes an analogous claim recognizing the demonstrative element in perceptual experiences, "there is a kind of thought we sometimes have, typically expressed in the form 'This F is G', and we may aim to have a thought of this kind when, in virtue of the absence of any appropriate object, there is no such thought to be had".

Although this view straightforwardly incorporates the particularity of objects, one problem is that this conception of content seems radically different to the one I introduced in Chapter 1 and in the preceding section. The content I talk about is something assessable in terms of veridicality, but it is hard to see how content as conceived of by Evans and McDowell can be wrong, given that their content is constituted by actual objects. When perception is nonveridical, the content must be different from the way the world is. For example, when you hallucinate a pink elephant in front of you, content that is assessable in terms of veridicality must reflect the fact that you experience an elephant despite its inexistence. Davies and McGinn's view is one such way. For them, perceptual content involves existential quantifiers and general terms. However, if content is object-involving, there is no content in the hallucination example because there is no object. It is therefore hard to see how conception of content can account for hallucinations.

Another related difficulty of approaches like Davies' and McGinn's is accounting for phenomenology. Such an approach assigns different contents to phenomenalogically indistinguishable perception and hallucination, because of the presence (or absence) of the object. Different contents will also be assigned to the pre-swap experience and post-swap experience: the content of the pre-swap experience involves (say) tomato_1 whereas that of the post-swap experience involves tomato_2. However, the tomatoes are subjectively indistinguishable, and so the idea that the two experiences have different content contradicts our phenomenology. Hence, this theory of content is not phenomenologically-apt[43], and so there is reason to think that a phenomenologically-apt theory would be better.

This approach rightly recognizes the role of demonstratives, however a problem arises from the supposition that singular terms have to succeed in reference to be contentful. If demonstratives can be contentful without building in their referents, the problem dissolves. Such an approach that can stay faithful to phenomenal particularity. Tyler Burge advocates a version of this position. He writes:

> [T]he Intentional content does not include any physical object that is actually picked out: the content is a demonstrative application of something of the form "That F is G". Sometimes a demonstrative content fails to pick out any objects. And sometimes, even when it does, it [the content] can be individuated independently of that object. The satisfaction conditions require that there be a relevant demonstrated object if the Intentional content is to be true (Burge, 1991, p. 196)

> To be *de re*, a thought should both contain a primitive demonstrative element in its content and involve successful reference through a demonstrative element to an object or *re* … [but] the two requirements for being a *de re* thought are separable. It is possible for an applied demonstrative element to

[43] Even though there are many difficulties with Evans' and McDowell's positions, many maneuvers have been made in the recent debates in philosophy of perception to rescue the spirit of Evans and McDowell's approach. See Soteriou, forthcoming.

> fail to have a referent. Since thoughts are individuated in terms of their contents, some demonstrative thoughts are not. Moreover, since some demonstrative token applications that in fact have a referent might have failed to have had one (if the contextual circumstances had been different) some thought tokens that are in fact *de re* are not essentially *de re* (Burge, 1991, p. 208).

It is important to emphasize the difference between Burge's position and that of Davies and McGinn. For Burge, phenomenal content involves singular elements along with general elements. This position is consistent with the swapping scenario. From Burges' perspective, in both pre- and post-swapping, the perceiver refers to the object by singular elements but the elements do not concern identity over time. We need not accept the generalist position of Davies and McGinn to account for the swapping scenario. But if singular terms can fail to refer, we can also build in singular elements into content.

Burge's account is arrived at as a response to the phenomenon of so-called veridical hallucination. A putative case of veridical hallucination is something like this:

> Veridical hallucination
> You are experiencing a red tomato in front of you. There happens to be a red tomato in front of you right exactly where and how you are experiencing, but your experience is caused by the effect of a tumor in your brain.

If you are with Davies and McGinn, this experience is veridical. Phenomenal content is characterized only with general terms and existential quantifiers, so the content in this case would be: there is a red tomato at location *L*. If that is so, this experience is completely veridical. But, most people have a strong intuition that the subject does not see the tomato; it is hallucination. Yet this is a "veridical hallucination". Burge's alternative is to say the putative case is just non-veridical. Perceptual content involves a demonstrative element and that element fails to refer in the case of veridical hallucination. For a demonstrative element to succeed in reference, according to Burge, the right kind of causal connection has to be in place, and this is absent in the hallucinatory case. Therefore, according to Burge, veridical hallucination (without quotation) does not exist.

Burge does not push his account as a way of accounting for the phenomenology of particulars, but Michelle Montague (2011) endorses a similar position aimed at explaining the phenomenological aspect of perception of particulars. Montague states:

My alternative account of the phenomenology of particularity begins with the proposal that many of our perceptual experiences (and thoughts) involve, as part of their basic structure, a bare demonstrative thought-form which can be represented as

[that (g) —]

where the blank is typically filled by ‘is F’, for some property F. (Montague, 2011)

Montague claims this is a kind of thought, whereas Burge is not committed to the claim. I will come back this issue in Section 6. What is important here is both Burge and Montague posit some fallible demonstrative element in perceptual content and Montague thinks that this explains phenomenal particularity.

Another prominent account of particularity in perceptual experience that avoids Evans' and McDowell's difficulty is Susanna Schellenberg's (2010). She thinks perceptual content is constituted by concepts that refer to objects and properties, but she also thinks content partly involves objects[44] Because she distinguishes between two kinds of particularity and seeks an account that can respect both kinds of particularity. One kind is phenomenological particularity, which corresponds to what I have been calling phenomenal particularity. The other kind is relational particularity, which is exhibited if and only if “the experiencing subject is perceptually related to the particular object perceived” (Schellenberg, 2010, p. 22).

It is because of this relational particularity that perceptual content in her account is object-involving. More concretely, she argues perceptual content can be properly understood as Fregean gappy content; the content is Fregean because it is constituted by modes of presentation. Schellenberg introduces modes of presentation to avoid

[44] It should be noted that her concept of concept is not a highly demanding one. She writes, “[A] subject who possesses the concept of redness has the ability to refer to red things, which involves discriminating red things from things that are not red.” (Schellenberg, 2010, p. 38)

metaphysical commitment to uninstantiated properties in the case of hallucination. If content is constituted by naked objects and properties, then in the casse of hallucination, there is no bearer of properties. Those properties are metaphysically problematic, so an account that can do without commitment to such properties is desirable. If modes of presentation are introduced, no such commitment is necessary because only concepts are required in order to have content. The relevant modes of presentation here are *de re* modes of presentation. The *de re* modes of presentation here are conceived as only partly object-dependent in the sense that there is at least some content even when no corresponding object exists. Thus, Schellenberg also admits perceptual experience can be contentful even in hallucinatory cases. However, the content is gappy where there is no object. The content can be described as follows. In the case of successful perception, the content can be described as,

$MOP_r(O_1)$

where MOP_r represents the *de re* mode of presentation and O_1 is the object of perception. If the experience is hallucinatory and there is no object, the content is,

$MOP_r(\_\_)$

In this case, the experience is constituted by the same mode of presentation but there is no object. Moreover, if the object is swapped with a numerically different but subjectively indistinguishable object O_2, the content is now,

$MOP_r(O_2)$

and importantly, $MOP_r(O_1) \neq MOP_r(O_2)$. Thus in Schellenberg's account, subjective indistinguishability is accounted for by the modes of presentation that are employed in both cases, and the relational particularity is accounted for by the potentially gappy content.

How does this account relate to the Burgean positions discussed earlier? Even though the demonstrative element is not explicitly referred to in this account, *de re* modes of presentation can easily be explained in terms of demonstratives. The serious difference only exists in the object-involving aspect, which was introduced to explain relational particularity and was not addressed in the Burgean positions. Thus whether Schellenberg's account is superior seems to hinge on the extent to which we should take relational particularity seriously. Burge-Montague's position seems superior as an account of *phenomenology* because it is an account of phenomenal content. However, even though content in Schellenberg's view is not phenomenal content (it is not characterized solely in terms of phenomenology), it does explain phenomenological aspect by modes of presentation in play. Therefore, I will suspend making a final judgement until later in the chapter. If it turns out that we should accept relational particularity, then Schellenberg's account is favorable. If not, Burgean positions are desirable. For convenience, however, hereafter in this chapter I will only talk about demonstrative elements and disregard modes of presentation and gappy content for convenience.

6.4. *Three kinds of phenomenal particularity and three layers of content*

In the last section, I discussed major attempts to account for particularity and came to certain degree of agreement with Burge's, Montague's, and Schellenberg's view. I will now look at what these views have to say about the swapping case.

Their views are sufficient to account for phenomenal particularity at a given time, but not when it comes to time-extended phenomenal particularity. Because of this, there are two possible ways to think about swapping cases. One way is to think both pre- and post-swapping experiences are veridical. Two different tokens of the same kind of perceptual demonstratives ("that tomato!") are in play. In this case, these demonstratives successfully pick up *different* tomatoes as they are only concerned with objects "here and now". I will call this notion of phenomenal particularity *narrow phenomenal particularity*. Another way to think about those experiences is to think there is one enduring demonstrative element in play. In this case, the post-swap experience falsely represents that "this tomato is the same as that one."

At least sometimes, we seem to track objects over time. For example, when you see a person, smiling and approaching you, but far away in a crowd, you might first not be able to identify her as your friend, Lisa. Later she comes close enough to you, and you finally recognize her. In this scenario, Lisa is not represented as Lisa at the initial stage; only at the later stage she is represented thusly. Yet, Lisa seems to be represented as the same object (or person) the whole time. I will call this kind of particularity *broad particularity*, particularity that is tracked over time. Narrow particularity and broad particularity are tightly related, as thin particularity is required for broad particularity. However, it seems to me that there are cases where identity though time is not required. For example, suppose you are at a café, trying to get a cappuccino to go. At the cashier, you find an empty cup, and recognize it. After a while a cup of cappuccino is served to you, and you go out of the café with it. In this scenario, you might not care if the cup is identical with the one you saw at the cashier. It may be so in reality, but you are noncommittal about it. This kind of situation seems realistic. Therefore either narrow particularity or broad particularity can be in play depending on the context and the conditions of the scenario. As it turns out, the mechanisms underlying phenomenal particularity are the ones that enable a subject to track an object *over time* by default, and how much longer the tracking survives depends on the conditions under which stimuli are presented. However, in the swapping scenario, the natural interpretation is to think broad particularity is involved, since subject would clearly judge that this tomato is the same as the one before (She would likely say, "What are you talking about? Nothing happened to it!").

It remains possible that some other form of particularity related to perceptual phenomenology exists, or that some kind of particularity always exhibited in perceptual experience? I think that both these possibilities will be answered in the negative. I will argue, however, that there is an aspect of phenomenology that does not exhibit any kind of particularity.

In one sense, it is hard to conceive of perceptual phenomenology that is not about the particularity of objects[45]. For example, Schellenberg (2010, p. 23) says, "Indeed it is unclear what it would be to have an experience that seems to be of a material, mind-independent object without it seeming to the subject that the object is a particular object." However, we can find the relevant aspect of phenomenology in the fringes of consciousness and in unattended aspects of consciousness. The fringe of consciousness is a difficult area to investigate introspectively, because every time we try to "look further," it comes under focus and so is no longer in the fringe. I argue that perception

[45] Phenomenology of mood might be another example that does not exhibit particularity, but since we are concerned with perceptual experience here, non-perceptual experiences are excluded.

in the fringe of consciousness is not about any particular object or objects.

In such cases, it can be useful to gain insight from cognitive science. In the psychological literature, a kind of perception called gist perception is widely studied. Gist perception is perception of the basic category of the whole visual scene as opposed to perception of specific objects in the scene. For example, when you are briefly presented with a view of a seashore, you probably only have a general impression that you have seen a scene of seashore without being able to recall all the objects in that scene (Oliva, 2005).[46] This implies that in the fringes of consciousness, we can perceive basic categories but not the particular items that comprise the scene. This point can also be illustrated by change blindness studies: if the change is congruent with the gist, it becomes harder to detect (Hochstein & Ahissar, 2002).[47]

The information about gist is processed quickly by a system that processes low resolution information (Kveraga, Ghuman, & Bar, 2007). This system produces an initial guess about the visual scene and initiates further informational processing to get more specific information. Studies about gist perception suggest the following view of the content of perception: the content of perception consists of the representation of gist and of particular objects. Gist perception is formed as a rough idea about the entire visual scene, with more detailed representations of attended objects congruent with the gist being formed afterwards.[48] We may experience a parasol and a few people in colorful swimsuits along with experience of gist of the beach, but we can never be able to experience all particular objects in the scene at the same time.

Taking all this into account, I conclude conscious gist perception does not exhibit either narrow or broad particularity. First, in gist perception, objects are not individuated. The representation of gist predicts the presence of certain kinds of objects, but it does not predict specific objects in specific locations. The prediction from gist representation about objects in the scene can only be characterized with existential

[46] You can see a demo on at https://www.k-state.edu/psych/vcl/basic-research/scene-gist.html.

[47] For example, a demo can be found at J. Kevin O'Regan's website: http://nivea.psycho.univ-paris5.fr/CBMovies/BigFishFlickerMovie.gif.

[48] Therefore, gist perception is not about the things perceived in the fringe. It is about the whole scene, and so the content of the representation of objects under attention is also congruent with the content of the gist. In the fringe of consciousness, there is no detailed representation of objects, so only gist is consciously perceived.

quantifiers and in general terms, just as generalists about content do. The representation is therefore along the lines of "There are objects that tend to exist on the beach". It is true that gist perception is also about some particular objects in the vicinity. Thus, gist perception is also not a weird kind of perception about abstract entities. My point here is that particularity is not reflected in *phenomenology*. Thee phenomenology of gist perception does not exhibit the "that thing!" aspect of phenomenal particularity. We have conceded that numerically different objects can give rise to the same kind of phenomenology. Thus particulars are not involved in phenomenal content, but there is phenomenological difference between when phenomenal particularity is exhibited and when it is not. In the former case, objects are individuated and demonstratively referred, whereas in the latter, they are not. Thus, we can say gist perception is a kind of perception without the phenomenology of particularity. It does not seem to the subject that any particular objects are presented. All of these kinds of particularity have to be explained. A taxonomy of particularity is given below (Table 4):

Table 4 Taxonomy of particularity

Narrow particularity	Perception that concerns particular objects at a point in time
Broad particularity	Perception that concerns the identity of particular objects through time
Non-particularity	Perception that does not concern any particular objects

It is often assumed that phenomenal particularity is an essential feature of perceptual content. I am inclined to think this is because philosophers have paid attention only to attended parts of perceptual experience. Nevertheless, it is worth considering some prominently presented reasons. Not much has been said on this topic, but I will discuss Sussana Schellenberg (2010), Matthew Soteriou (2000), Tyler Burge (2010), and John Campbell (2000).

I will begin with Schellenberg. She explains why particularity should be incorporated as follows:

The motivation for this way of understanding accuracy conditions is that the condition that needs to be met for an experience to be accurate is not just that there is an item in the world that possesses the properties specified by the content. It is necessary to specify which particular object in a subject's environment is represented to determine whether the subject's environment really is as it is represented to be. (Schellenberg, 2010, p. 21)

Soteriou uses similar reasoning. He also thinks content has to be particular-involving to enable veridicality to be assessed in the right way. His interest is in illusory cases. In illusory cases, the subject does perceive an object, but gets some of the object's properties wrong. In such cases, it is fairly safe to say the experience is not veridical. If illusion is possible and the experience is non-veridical, we need some way to identify which object about the perception is about.

What should we think about it? Does this point necessitate particular-involving content? First of all, particulars *themselves* need not be involved for the assessment of veridicality discussed above. Singular *representation* of particulars (i.e. demonstratives) seems enough to do the job for the following reason. Recall that Burge argues that socalled veridical hallucination is impossible; hallucination just is a non-veridical case. This is because in the hallucination case, singular representation is not invoked through the right kind of functioning of visual systems. According to Burge, perception is about the entities in the world that cause perception in the right way. The entities have to invoke the right functioning of visual systems (Burge, 2010, p. 381-2). Extending this reasoning to other cases, the relevant object for a singular representation is the one that invoked the representation through normal functioning of the visual system. For example, a singular representation of a cup can be about the cup to my left, when it is caused via the following causal chain: a light reflected from a cup to my left hits the retina of both eyes and those sensory signals are sent up to visual primary cortex via lateral geniculate nucleus and so on.

Secondly, the representation of gist seems to require separate treatment for veridical conditions. The experience of gist does not exhibit particularity, and representation of gist does not involve the representation of objects. Can gist perception be non-veridical? If so, how is its veridical condition given? To answer these questions, we have to answer ontological questions about gist.

Whether some scene falls under a scene category is usually not decided on the basis of a*n* object, but decided on the basis of a *collection* of objects. We can say the category of a scene supervenes on the arrangement of objects in the scene. For example, a scene becomes a scene of a seashore by virtue of the existence of characteristic objects: the sea, sand, palm trees, and so on. The relationship between the arrangement of objects and scene category is complex, as the concept of scene category is a *cluster concept*; there is a weighted list of criteria to fall under the category, with no item being either necessary nor sufficient. For example, palm trees tend to exist on the seashore, but they do not need to be nor they are sufficient for the scene to be the seashore.

Given these points, we can decide about the veridicality of gist perception in reference to collections of objects in the scene rather than a particular object. There could be a scene category, in which one item has more weight in deciding whether the scene fall under the category.[49] However, even in those cases where only a few items are important, what matters is that all objects exist *within a certain space*. In this sense, scene category is defined in terms of space. Thus, in answer to the first question, we can suffer from illusions of gist. There can be cases where there actually is a town when it seems to us there is a forest.[50]

Another argument to show the necessity of particularity concerns the function of perception. Burge argues that representational content involves particular objects because "the practical function of perception is to enable individuals to engage successfully with the particulars in their environment" (Burge, 2010, p. 381), and our perceptual system is designed to represent particulars in our vicinity for that purpose. If that is correct, does perceptual content always have to involve singular terms? I think the answer is negative. It is true that we deal with particulars in our vicinity for survival, but as I will discuss later, our ability to track objects is limited; we can only experience up to five objects at the same time. We frequently experience more at a given time than our capacity for experiencing objects *as* objects allows. It is natural to think that kind of

[49] How many objects are involved in determining a scene category also is the matter of how to individuate objects. I have been talking of objects at an intuitive level, but we can talk of collection of dancers as an object, a person as an object, or someone's right hand as an object.

[50] There could also be hallucinations of gist. If the subject suffers from hallucination of objects, then there seems no reasons to think hallucination of gist is impossible. Whether hallucination of gist can be brought about without any representation of objects and their properties is an interesting question.

experience—namely experience of gist perception—also performs some practical function. Gist perception provides a "general summary" of the surroundings and by doing so helps the subject navigate the objects around her. It provides contextual information that can be used to predict what kinds of objects we can further expect under the circumstances.

There would also be an epistemological worry. John Campbell (2009) thinks conscious perceptual experience enables us to have conceptual thoughts about external objects and properties. Conscious awareness of a chair enables us to have the conscious thought that that chair is red, for example. If that is so, one might suspect that gist perception cannot play any epistemological role in our cognitive life because gist perception does not involve any object individuation. Regarding this worry, I have to admit that gist perception does not enable thought whose content takes the form that S is F. However, as discussed above, the content of gist perception can rightly be described using existential quantifiers, and thus it can enable thoughts with that kind of content. For example, we can k, on the basis of gist, that there are some kitchen tools (given the scene is of a kitchen).

So far, there is no conclusive reason to think all perceptual content should involve singular elements. Far from that, we found some positive reasons to have content that does not involve singular elements. First, we found a practical function of gist perception, which is different from perception involving singular elements. Second, we found types of thoughts that gist perception enables. This motivates us to think gist enters the content of perception. Why has this aspect of experience overlooked by philosophers? As I briefly mentioned above, I think this relates to a methodological problem with introspection. Gist perception is an unattended aspect of perceptual experience. If philosophers who theorize about the content of conscious perception rely on introspection and introspection requires attention, it is logically impossible to investigate unattended aspects of conscious perceptual experience.

In this first part of the chapter, I discussed the phenomenology of perceiving particular objects. I agreed with Burge and Montague that the content of perceptual experience that exhibits particularity is characterized by demonstrative elements. I also proposed two kinds of particularity and an aspect of phenomenology that does not involve particularity at all.

6.5. *Particular object representation as a problem for predictive processing*

In the second part of the chapter, I discuss mechanisms that underlie phenomenal particularity and give a predictive processing account for that.

At a glance, it looks as though phenomenal particularity calls for inferential mechanisms, but influential accounts for perceptual demonstrative mechanisms, visual index theory and object file theory, are non-inferential views. Indeed, one of major proponents of the accounts, Xenon Pylyshyn (2009), gives an objection to inferential accounts. The objection can be answered when the commitment of predictive processing is reconfirmed: predictive processing does not entail coherence between perceptual content and the content of beliefs.

Let us first look at visual index theory and object file theory. Visual index theory addresses the initial individuation of objects and maintenance of object representation in a visual scene. According to the theory, the initial individuation of objects is achieved in the manner of demonstrative reference (Pylyshyn, 2001, p. 129). According to Pylyshyn (2001), the functioning of visual indexes relies on the early vision system, which does not require any explicit representation of the properties of an object and functions in a modular way. One important strand of research supporting this theory is the multiple object-tracking paradigm conducted by Pylyshyn and his collegues. In the experiment/study, subjects are asked to track otherwise identical objects positioned differently in a display. The results indicate that subjects can only successfully track up to five objects at the same time (Figure 11). This suggests that the brain is equipped with an apparatus that works independently to property representations (except for shape and location). In reference to our taxonomy of particularity, we can understand narrow particularity in terms of the initial assignment of demonstratives (what Pylyshyn calls “indexes”) to objects, and broad particularity can be understood in terms of the maintenance of the assigned demonstratives and property attachment to the demonstratives, which will be explained shortly.

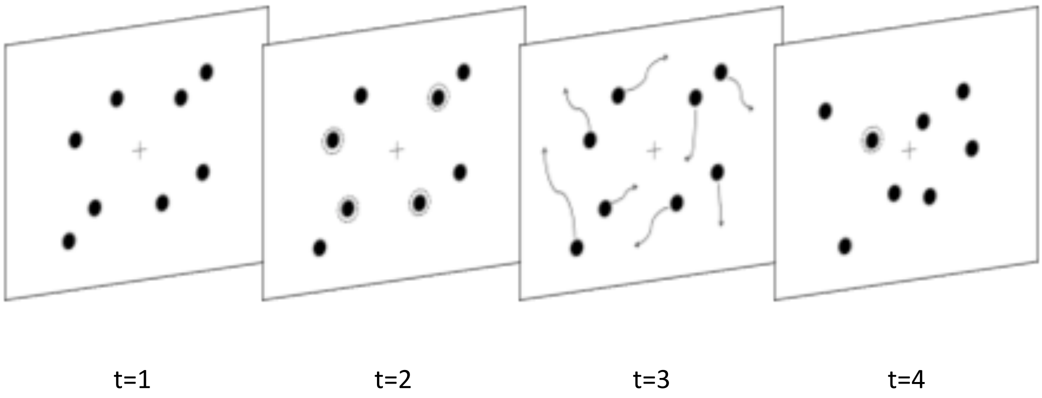


Figure 11 multiple object tracking paradigm

According to visual index theory and its sibling theory, object file theory, proposed by Daniel Kahneman and Anne Triesman (1992), the singular representations ("indexes") first emerge as bare demonstratives: demonstratives without any property attributions. Properties are incrementally attached to the representation in later stages and the representation of the object and properties are called the object file. The gradual build-up of the singular representation can easily be tied to broad particularity, as broad particularity is a matter of representing a particular object as something that has existed over time. According to object file theory, the recognition of continuity or discontinuity of objects takes three kinds of operation: (1) a correspondence operation, (2) a reviewing process, and (3) an impletion process (Kahneman, Treisman, & Gibbs, 1992). The correspondence operation determines if an object in view is a novel object or athesame object with different spatial property. A reviewing process retrieves information about the initial object, and an impletion process "bridges" the gap between the current object and the past object by the using current information and the past information, yielding a percept of change or movement of objects. They refer to a phenomenon called apparent motion as an example. In apparent motion, the first stimulus is briefly presented and taken out. After an interval, another stimulus is presented at a different location. If the spatiotemporal difference between the two stimuli (the location of disappearance and appearance and the length of time interval) is within a certain range, the subject perceives smooth motion.[51] These parameters have to be within an ecologically plausible range, otherwise the subject experience a discontinuous disappearance and appearance. This judgment about the novelty or sameness of an object is made only on the basis of spatiotemporal contiguity (Kahneman et al., 1992, p. 180). Thus a change from a frog to a prince can be perceived as a change of one object, rather than two distinct objects if the locations of those two objects are close enough and the interval is short enough.

In contrast to the initial individuation of objects, the object file stores representation of properties that belong to objects. The properties attached to the object file would vary. They can be relatively high-level properties such the property of being a frog or a prince—or they can be as low as spatial properties or color properties. This is also supported by the recently proposed neural object file theory (Xu & Chun, 2009). Xu and Chun investigated areas of the brain implicated in these processes and found involvement of the superior intraparietal sulcus and higher visual areas. This underwrites the potential participation of high level properties to object files.

These theories and their evidence might look congenial to inferential views such as predictive processing. In the case of apparent motion, the motion is perceived only when the spatiotemporal difference between the two stimuli is psychologically valid. It looks as though after the brain takes in the information regarding the second stimulus, it judges there has been a motion in a backward way. Another example is the multiple object-tracking paradigm. What is particularly relevant here is that tracking is sometimes successful even when the stimulus is not visible for short duration; an object can be tracked even when it is occluded by another object. Another interesting feature is that an object is not tracked when the object dwindles and disappears at a point and another object appears and looms from the same point (Pylyshyn, 2001). In the latter case, despite its spatial contiguity, the two objects are represented as different objects. Again, this seems a case for perceptual inference. The brain makes an inference based on its belief about the endurance of objects, namely the belief that an object can still exist even when it is occluded by another object and that once an object is diminished, it will not be restored.

[51] A demo of apparent motion can be found at http://www.michaelbach.de/ot/col-colorPhi/

Despite the *prima facie* case, there is a problem: the inference is not responsive to the contents of your conscious beliefs. The two dots in apparent motion are *not* actually moving (hence apparent) and you *know* it. They are two different dots and you know it. Despite your conscious knowledge, your brain keeps producing the percept of motion! Perhaps not all types of perception, but most types of perception are incorrigible by conscious beliefs.

This is more vividly illustrated by the example from Carey and Xu (Carey & Xu, 2001). Look at Figure 12 below. If you are first presented with Panel 1, deprived of that, and then presented with Panel 2, you would judge two creatures, the pig and the duck made a diagonal jump across the panel, with the pig moving from the upper left to the bottom right corner, and the duck moving from the bottom left to the top right. However, according to Carey and Xu, if Panel 1 and Panel 2 are presented successively (i.e. with no time gap) and the conditions are arranged so that apparent motion takes place, you would have a surprising perception. You would perceive two creatures jumping horizontally and metamorphosing from a pig/duck and into a duck/pig! In the former case, according to Carey and Xu, you make a judgment bearing on concepts. You have the concepts of rabbit and bird and you believe it is very unlikely that a rabbit is changing into a bird (or vice versa), that is why you judge the rabbit and the bird moved diagonally. In the latter case, your visual brain unconsciously "judges" that there are two very surprising creatures in front of you despite your conscious judgment to the contrary. This is because the system is run almost solely based on spatial relationships and the distance between two upper (or lower) corners is shorter than that of diagonal line. Thus, Carey and Xu argue there are two systems that produce the representation of particulars and the contents of the representations can conflict with each other. The visual system responsible for the representation of particulars seems to care only about specific kinds of information, and this content cannot be informed by your conscious beliefs.

These examples, at minimum, show that perceptual processing is procedurally and qualitatively different from conscious inference. The contents of beliefs tend to be coherent, they can in principle interact each other, and inference processes can be conscious even if they are not always conscious. On the other hand, perceptual content can be incompatible with the content of beliefs, and is in most cases incorrigible. Additionally, the inference processes are almost always not accessible to consciousness.

If so, in what sense can perceptual processings be said to be inferential (or inference-like)? How can predictive processing explain all these phenomena? In Section 7, I will propose predictive processsing account to answer those questions but before that, I will develop the argument against inference view given by Pylyshyn.

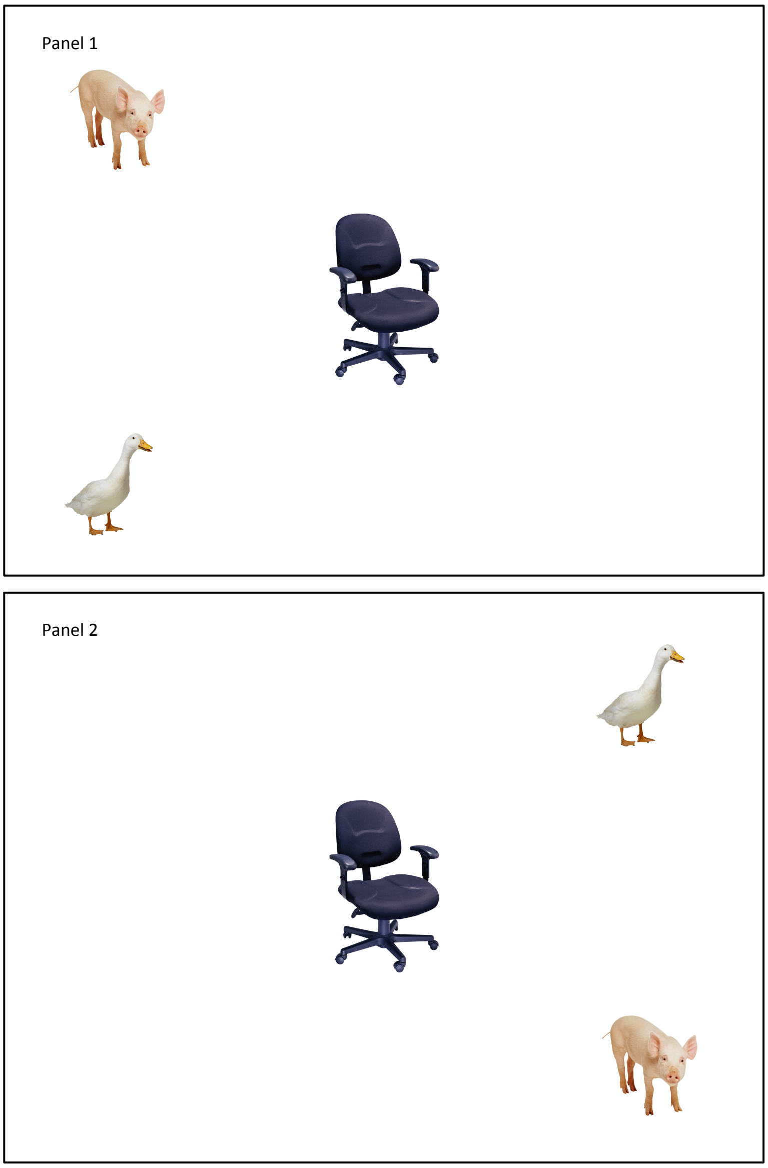


Figure 12 a Carey and Xu type of example which leads to the two-system view

6.6. *The noninferential nature of the early visual mechanism*

The problem of insensitivity to other beliefs is not confined to cases involving particulars. Rather, the problem happens in many instances of perception. In this section, I firstly discuss the argument against perceptual inference given by Pylyshyn (2001). I will then discuss how the problem specifically relates to the representation of particulars.

The idea of perceptual inference in general is heavily criticized by one of the main investigators of the representation of particulars in cognitive science, Xenon Pylyshyn. He claims that there is, at least, a visual mechanism that does not involve any kind of inference (Pylyshyn, 1999). He calls this mechanism the *early visual mechanism*, after David Marr, and examines its properties. According to Pylyshyn, the computations in the early visual mechanism are better conceived of as embodying constraints from nature;the mechanism is not performing any inference at all. One main thrust to advance the idea of perceptual inference was the inverse problem from sensory stimulation to worldly causes, as discussed in Chapter 2. It is argued that the brain has to "infer" based on prior beliefs about how the world is. The natural constraint is an alternative mechanism to overcome this inverse problem. As we will see shortly, natural constraints play a role in narrowing down possible sensory causes by making some assumptions about them. By virtue of the constraints, representations are veridical in most, if not all, situations.

One such example discussed by Pylyshyn is the "rigidity" principle studied by Ullman. The principle states that "if a set of points moves in a way that is consistent with the interpretation that they lie on the surface of a rigid body, then they will be so perceived" (Pylyshyn, 1999, p. 354). Since points that move in that way tend to be actually on the same surface, that kind of perceptual representation would most likely be veridical. This also explains our illusory percept of a surface as that of a kinetic depth effect. In kinetic depth effect, subjects are shown dots moving on a computer screen, but because the movement of the dots is consistent with the interpretation that they are on the surface of rigid objects, subjects perceive invisible surfaces of 3D-objects. The rigidity principle explains and predicts when and why that kind of illusion occurs. According to Pylyshyn, phenomena that are more associated with inference can also be explained in terms of the idea of the natural constraint. Perceptual constancies are phenomena where the perception of color, size, brightness, and other propertiews, seems to take a variety of environmental factors into account. For example, a part of a white paper usually continues to look white even when it is shadowed. It seems as though the brain infers that the color of the shadowed part of paper is white based on the knowledge of the difference in lighting and the knowledge that the paper likely has a uniform color. But, Pylyshyn argues this kind of phenomenon should also be construed of as computations that embody natural constraints, such as “luminance differences [that] are caused by reflectance-property differences or by illumination differences” or “that illumination tends to be equal for nearby regions in a plane”, for example. Why should the natural constraint view be preferred over the perceptual inference view? It can be objected that the phenomenon of moving dots can equally be explained in terms of perceptual inference, because the relevant visual system infers the existence and motion of dots in space from the prior belief that dots moving in a particular way lie on the surface of a rigid object.

Pylyshyn raises a few reasons why those computations are not inferences. One reason is the inaccessibility, to cognitive or other mechanisms, of the contents of the natural constraints embodied in early vision. The principles discussed above that enable specific kinds of perception are discovered as the results of empirical inquiry and not introspectively available to consciousness (and thus they are embodied by computations, not represented).

Another reason is the impenetrability of computations in the early vision system. These computations are not susceptible to influence from any other kind of knowledge. In the case of the moving dots, even when we know there are only dots on the screen, the dots produce the percept of invisible rigid objects with dots moving on their surface. This suggests that knowledge cannot influence computation inside the early vision system. Thus the contents of perception can cut across the contents of cognition.

In other words, Pylyshyn thinks that to count computations as inferences, the contents of representational states involved with perceptual inference would have to show some integration with the contents of other representations such as beliefs. Pylyshyn concludes as follows:

> Terms such as “knowledge,” “belief,” “goal,” and “inference” give us an explanatory advantage when it allows generalizations to be captured under common principles such as rationality or even something roughly like semantic coherence. In the absence of such overarching principles, Occam’s Razor or Lloyd Morgan’s Canon dictates that the simpler or lower-level hypothesis […] is preferred. (Pylyshyn, 1999, p. 357)

We can summarize Pylyshyn’s argument against perceptual inference as follows:

1. If the early visual mechanism is inferential, it has to show semantic coherence with conscious beliefs.
2. The early visual mechanism does not show semantic coherence.
3. Therefore, the early visual mechanism is not inferential.

In the next section, I give a rejoinder from predictive processing perspective rejecting the first premise.

6.7. *A predictive processing account of the representation of particulars*

In this section, I reply to Pylyshyn's argument and provide a predictive processing account of the representation of particular objects. A predictive processing account can reject the first premise of his argument. Remember *Prediction Error Minimization, Cognitive Penetration,* and *Perception in Predictive Processing* from Chapter 2. What makes up perception is what is used to explain away sensory stimulation, but when we are engaging conscious thinking, models are not used to explain away sensations. Therefore, conflicting models of perception and thinking are not employed at the same time. As explained in *Cognitive Penetration*, top-down influence from thought, like high-level representation, is expected when precision is low, but the cases under discussion are not such situations. Thus, the basic explanation is the same as the one for illusion: there is no prediction error being explained by high-level representations. Recall the explanation of Müller-Lyer illusion in Chapter 2. Basically, illusion takes place because the illusory percept was the best guess from the sensations experienced. Because of that, there is no prediction error sent to higher-levels.

In the previous section, Pylyshyn argued that visual processes are not inferential and that they only embody natural constraints. It was called natural constraints because the behaviors of visual mechanisms are not malleable. However, in terms of a predictive processing account, it can be argued that he models simply suppress prediction errors. Thus, natural constraints can be understood as constraints that have to be satisfied by models that best suppresses prediction errors.

In a way, it is all too obvious that perceptual inference, taken at face value, is not inference. What is important here is sense in which the processes are inference-like. In Chapter 2, I made a terminological caveat that cognitive terminology should not be taken in a literal way. Instead, I concluded that predictive processing can account for the early visual system and its independence from conscious beliefs. In the next section, I will focus specifically on the representation of particulars.

6.7.1. Incorporating the visual index mechanism into predictive processing

The previous section was focused on vision in general, but we can tell a similar story about the representation of particulars. Apparent motion and multiple object tracking corresponds to the Müller-Lyer illusion. As discussed, apparent motion has both an aspect that is compatible with perceptual inference and an aspect that is incompatible with perceptual inference. On the one hand, apparent motion is in a way illusory, and this is known to the subject. At the same time, the processes that lead to the formation of the percept are largely unknown to the subject. On the other hand, apparent motion can be explained in a parallel way as the product of local perceptual inference. The brain behaves as if it knows how an object moves in the world: movement takes place only under "ecologically valid" conditions. Indeed, Kahneman describes the phenomenon in the following way by object file terms:

> If a physically plausible displacement or transformation could result in a match [between the first and the second stimulus], the relevant object file may be updated and the transformation may be seen to occur […]. However, if the new stimulus is sufficiently different from all its predecessors, or if the change in location is incompatible with the time interval or with any previous trajectory, a new object file may be opened and the sudden appearance of a new object will be consciously experienced. (Kahneman et al., 1992, p. 179)

Thus, the conscious percept of motion seems created based on the brain's "knowledge" about movement in the world.

We can adjudicate both aspects by explaining this as a local inference that takes place at lower levels. The inference is only a "local" one as the inference processes are immune from the influence of higher-level beliefs; the perception cannot be ceased no matter how strongly you (correctly) believe that there is no actual motion. Similarly to the case of the Müller-Lyer illusion, the model can explain incoming signals so well that no prediction error is sent up to the higher-level models.

We can apply parallel reasoning to multiple-object tracking too. Multiple object tracking also has the two aspects just discussed. The assignment and maintenance of indexes is performed largely independently from what you consciously think. However, the brain seems to have performed some inference to generate the illusion. As previously mentioned, the brain seems to know how an object behaves in the world. The index of one object survives occlusion by another object, and a new index is created when an object appears in place of a perished one. Even at the initial assignment of indexes, we can find some inferential elements. Pylyshyn writes, "[E]arly visual processes segment into feature-clusters which tend to be reliable proximal counterparts of distinct individual objects in a distal scene" (Pylyshyn, 2001, p. 146). Here, the brain seems to make an inference from some feature clusters to objecthood. All these elements can again be treated as a kind of local inference. They sit at lower levels and suppress prediction errors so well that no prediction errors are sent to higher levels.

There are cases where these lower-level local inferences are not strong enough to resolve objecthood. In those cases, the higher-level models are called into play. The example Carey and Xu's example, described above, is one such example. The case where apparent motion does not take place is precisely the case where lower level models are not good enough to suppress prediction error. The models were not good enough because the spatiotemporal difference between two stimuli was not ecologically valid. Therefore higher-level models that involve information about the category and properties of objects are called into work. Carey and Xu needed two systems to explain the phenomena, but predictive processing can explain this within a single framework—they are just related to models operating at different levels in the hierarchy.

6.7.2. Solving the problems from predictive processing perspective

In the previous section, I replied to Pylyshyn's argument against perceptual inference, and gave a predictive processing account of the representation of particulars. Before closing this chapter, I will address the three problems other than the description problem discussed in section 4. The remaining problems are: the perceptual problem, the inverse problem, and the reason problem. I will address them in that order.

Regarding the perceptuality problem, Michelle Montague (2011) argues that the phenomenology of particularity is actually perceptual and also that conceptual thoughts underlie the phenomenology. She thinks a kind of demonstrative though is responsible for the phenomenology of particularity and that phenomenal particularity is therefore a kind of "cognitive phenomenology". I am sympathetic to her analysis in that she also addresses the demonstrative aspect of phenomenal particularity, but I do not think full-fledged conceptual thoughts underlie phenomenal particularity.

To start with, demonstrative representations are undoubtedly part of perception as they play a pivotal role in explaining away sensory stimulation (cf. Perception in Predictive Processing in Chapter 2). Thus, the problem is whether demonstrative representations are more percept-like or concept-like. Given what we saw in our discussion of visual index theory and object file theory, demonstrative representations themselves are quite primitive within the visual system. The conceptuality of the representations depends on what properties are attached to them. Firstly, Xu and Chun's (2009) neural object file theory demonstrates that the object file can involve the representation of higher-level properties and the activation of higher-level visual areas. Secondly, higher-level properties can involve object identification when other information is not available, as in the case in Carey and Xu (2001). Interestingly, under *Perception in Predictive Processing*, both of the cases in Carey and Xu can be considered perceptual. Even in the case of the diagonal jump, the resultant percept is the outcome of prediction error minimization (if conscious thinking is not involved).

Regarding the inverse problem, we have already seen two ways to solve the inverse problem: the natural constraint view and predictive processing. The sensory stimulation is indeed ambiguous between many possibilities—the hypothesis that there are two separate dots and the hypothesis that there is one dot moving, for example. The natural constraint view solves this problem by positing an assumption about the world, but the predictive processing view solves it by informative priors about the world. Both views are compatible each other, but the predictive processing view may be said to be superior as it predicts when more top-down processes matter (i.e. when the estimated precision is low). Also be said that because of unification virtue, which I discussed in Chapter 1, the predictive processing view is superior because it explains many diverse perceptual phenomena via the single Bayesian mechanism.

The answer to the reason problem is also fairly simple. As I argued in Section 4 (following Burge), the primary function of perception is to enable the perceiver to navigate particulars in the environment., Hierarchical predictive processing mechanisms enable this by deciding when to rely on shorter timescale causal regularity, and when to use longer timescale causal regularity. When information about shorter timescale causal regularity is not sufficient to decide particular objecthood, the brain resorts to longer timescale causal regularity (such as the case of Carey and Xu's example). By doing this, particularity becomes just another inference problem that hierarchical predictive processing can solve.

6.8. *Conclusion*

In this chapter, I tried to explain the representation of particulars from the perspective of predictive processing. We encountered the problem of unresponsiveness to conscious beliefs. The explanation for this phenomenon is that this is not a problem for the predictive processing framework since in such cases, there isno prediction errors being passed up the chain. To reach this conclusion, we revisited the idea that "inference" in predictive processing is not inference taken at face value, and should be understood as a technical term. Thus the framework well explains how phenomenal particularity takes place and when higher level models matter. One might think this is just a reinterpretation of the natural constraints view; a relabeling using predictive processing terms to avoid empirical appeals. I think this is a fair concern. Perhaps, the predictive processing account here does not make any different empirical predictions to the natural constraints view. However, as I noted above, there are benefits to *unification*. Predictive processing is a powerful theory that is expected to explain the functioning of the whole brain, and my attempt is to explain various perceptual phenomena from its perspective. The predictive processing account offers the benefit that the representation of particulars can also be explained in the same framework that can explain many other mental phenomena (including four other Ps).

7. *The fifth P: Persons*

7.1. *Introduction*

In the 2000 movie *Cast Away*, the main character, Chuck Noland (played by Tom Hanks), is stranded on a desert island due to a plane crash. He manages to survive on the island all by himself, but is in dire need of another person to talk to. In the more classical castaway story, Robinson Crusoe was lucky enough to find a servant from among the prisoners of the Cannibals, but Chuck was not lucky enough to meet any real people. He ends up drawing eyes and a mouth on a volleyball with his own blood and names the ball Wilson. He talks to the ball, and it becomes irreplaceable company given his isolated life on the island. Indeed, one of the most dramatic scenes of the movie was when Wilson fell off from the raft Chuck was rowing to escape from the island. Wilson could do nothing but float (of course) and it was evident that Chuck should remain on the raft, and had to abandon Wilson. Chuck was crying and apologizing to the floating volleyball. For Chuck, this object was not just a ball, it was Wilson: his best friend. This movie, though fictitious, shows the indispensability of other persons, especially ones we are close to, in life.

Getting food, sharing the road with other drivers, riding the train, and all sorts of daily activities require interaction with other people, and this this requires some kind of recognition of them. For example, it is certainly important to recognize a person as person, and often as a specific *type* of person. When you want to pay the bill at a restaurant you would first need to recognize who the persons in the restaurant are; you have to distinguish them from other objects such as plates, table, cash register, and so on. (Intuitively, this seems all too easy, but from the perspective of a cognitive system, or artificial intelligence, it is not). Then, you need to identify a restaurant employee from among those persons. In this case, it does not matter who the restaurant staff member is; individual difference does not matter, as long as the chosen person is a restaurant staff member, it is enough. However, in some cases, particularity matters. If one of your friends asks to borrow some money, you would want to remember whether she has returned previous loans to you within a reasonable timeframe. This requires you to have the ability to represent a particular person with their historical properties.

This capacity is easily evidenced among linguistic creatures like us, but is probably not limited to human beings. Identifying an individual and its history seems highly important for the reproductive success of any creature of a social species. In the discussion of the evolution of reciprocal altruism, some evolutionary biologists have claimed that the existence of cheater detection systems is necessary for the evolution of reciprocal behavior (Cosmides & Tooby, 1992). That is, an animal has to be able to detect another member of their species who hasn't cooperated in the past. A famous putative example in animal kingdom is vampire bats, who live entirely on blood. They die after just two days without food, so sharing blood with other members of their species is a rational strategy. It has been reported that vampire bat feeds other members of their species, but only those from whom they can predict future reciprocation (Carter & Wilkinson, 2013).

This aspect of experience is my focus in this chapter: the identification of familiar persons. I will mostly focus on cases where humans recognize other human beings, because those cases are the best studied. However, this form or perception is not limited to humans, as demonstrated by the above example. There are also cases where humans represent particular ordinary objects (e.g. a watch) or even places as special, familiar ones. Thus in the most general form, the topic of this chapter is the recognition of entities with which the experiencer has a personal interaction history and knowledge obtained through those i interactions. This is closely related to topics in the chapter on Particularity. There I discussed issues related to the individuation of objects, online tracking of objects though time, and identification where conceptual knowledge also matters. But the cognition discussed here is even more "intellectual" in the sense that personal interaction history and knowledge is required. Because of this, this aspect can be said to be an even deeper aspect of the world, as it is not clear how we can represent personhood (or its more general forms) without relying on prior knowledge about the entities involved.

I will develop my account mostly by way of explaining cases of *dysfunction* of the identification of persons—misidentification delusions. The breakdown of the system sheds light on the mechanisms involved in identification of persons. Thus accounts of personal identification were developed hand in hand with studies about identification dysfunction.

The chapter proceeds as follows. I will start by considering existing accounts of misidentification delusions. The most popular account of misidentification syndrome pins down perceptual and cognitive problems in patients. Though consistent with predictive processing, it lacks a concrete functional account of the perceptual disruption. It suggests that a disruption in the autonomic system causes a problem in the perceptual identification of a person, but it lacks the story of how it comes about. The mindreading account suggested by William Hirstein will be discussed as a candidate to bridge the gap. However, thorough consideration of the problem from a Bayesian perspective leads us to a more sensory account. In the final section, I suggest a sensory signature theory of person identification.

Alongside giving this account, I will answer four questions raised in Chapter 1, namely, the inverse problem, the description problem, the perceptuality problem, and the reason problem. As a refresher, the inverse problem is about how we can experience personal identity given that our sensory stimulation is limited, the description problem is about how to describe the phenomenology of persons aptly, the reason problem is about why we experience such a deep aspect of the world—personal identity. Finally, the perceptuality problem is about whether our experience of persons is perceptual or not—in some cases, identification seems perceptual: you instantly recognize who a person is. When you see the face of a close family member, you would instantly recognize their face. However, in other cases, identification is a more elaborate attempt. At a party, you might encounter someone you only met somewhere once before; the person seems to know you, but you cannot identify the person instantly. In those cases, you would try to carefully figure out who he or she is through conversation. These two modes of personal identification will be addressed. All of the four problems will be discussed in the final section.

7.2. *Misidentification syndromes*

Misidentification syndrome is an umbrella term for delusion disorders that involve the misidentification of a person. Delusions themselves are usually characterized as "fixed beliefs that are not amenable to change in light of conflicting evidence" (American Psychiatric Association, 2013). Their contents may come in many varieties. Persecutory or referential delusions are likely the most well-known. An example of a persecutory delusion is a case where the patient believes that she is being chased by CIA agents. An example of a referential delusion is a case where the patient believes that she is being referred to by every passer-by. Patients with delusions keep holding the delusional beliefs despite the testimony of doctors, surrounding people, or evidence to the contrary.

The relevant kinds of delusion here are delusions that involve the misrepresentation of a person. In misidentification syndrome, a person (or multiple persons) are taken to be a different person (or persons) to who they really are. There are different versions of misidentification delusions (see Table 5 below), but what I will focus on here is the Capgras delusion and the Fregoli delusion, because both are well documented and studied. The Capgras delusion and the Fregoli delusion are, in a way, twins. The Capgras delusion involves the misidentification of a familiar person as a stranger, whereas the Fregoli delusion involves the misidentification of a stranger as a familiar person in disguise. In Capgras delusion, one believes that a person who is emotionally close to her has been replaced by an imposter. Even though the appearance of the close person has not changed, the Capgras patient insists that it is an imposter who just looks like her. On the other hand, the Fregoli delusion is usually characterized as "the mistaken belief that some person currently present in the deluded person's environment (typically a stranger) is a familiar person in disguise" . In the original case of the Fregoli delusion, a young woman in Paris firmly believed that two actresses in Paris, Sarah Bernhardt and Robine, had been chasing her.[52] According to the patient, these actresses fit their appearances into those of her friends or whoever she might know, "including the strangers she saw in the street, doctors, friends, and previous employers" (Langdon et al., 2014, p. 616). Thus, in this case, even though the appearances of the targets of her delusion were diverse, the patient kept believing that they were all either Sarah Bernhardt or Robine. As is clear in this case, the target of the Fregoli delusion does not have someone whom the patient knows personally. This particular patient is reported to have been a theatre enthusiast, and Sarah Bernhardt and Robine were frequently seen by her. Thus, she may have feeling of familiarity toward them. In both delusions, obvious similarity (or lack thereof) in appearance is not taken into account (or is disregarded after consideration) in the making of judgments about personal identity.

[52] This delusion is named after the Italian actor and mimic Leopoldo Fregoli who is known for his impersonation skills.

Table 5 Varieties of misidentification syndrome

Misidentification Disorder	**Content**	**Literature**
Capgras delusion	A person who is emotionally close to the patient is thought to have been replaced by someone different.	Capgras and Reboul-Lachaux (1923), translated by Ellis et al. (1994)
Fregoli delusion	Some person currently present in the patient's environment is thought to be a familiar person in disguise.	Courbon and Fail (1927) de Pauw et al. (1987)
Mirrored self-misidentification	One's self reflected in a mirror is identified as a stranger.	Breen, Caine, and Coltheart (2001)
Intermetamorphosis	A known person is thought to have been transformed into another known person, who is psychologically and physically different.	Courbon and Tusques (1932), Ellis et al. (1994)
Reverse intermetamorphosis	The patient transforms him/herself into another person	Breen, Caine, Coltheart, Hendy, and Roberts (2000)
Subjective double	Another person is thought to be a duplicate of oneself	Christodoulou (1978)

Another disorder often discussed in the literature, while not a misidentification delusion, is prosopagnosia. Prosopagnosia is disorder of facial processing (Ellis & Young, 1990). A patient with prosopagnosia cannot recognize familiar faces despite preserved visual processing. This disorder is interesting because the ability to perceive faces seems selectively impaired. It is usually associated with fusiform gyrus, which seems to be a selective responsive to face stimuli, and is impaired in prosopagnosia. There are two types of prosopagnosia, apperceptive prosopagnosia and associative prosopagnosia. Patients with apperceptive prosopagnosia cannot make sense of faces at all. They instead see faces as blobs or caricatures. They cannot perform same/different judgments with different faces, nor can they differentiate familiar and unfamiliar faces. In associative prosopagnosia, patients have more spared perceptual capacities: they can correctly make same/different judgments and can correctly guess sex and approximate age and a picture of a face. Thus, they seem to be able to perceive faces as faces to some degree. However, they cannot identify faces as the faces of specific people. Thus associative prosopagnosia is the disorder more relevant to the problem of Persons. Interestingly, there is a subset of prosopagnosia patients who show covert responses to familiar faces even though they cannot explicitly recognize them. For example, some prosopagnosic patients can correctly guess identity of faces while their confidence on their own judgment is low (Rivolta, Palermo, Schmalzl, & Coltheart, 2012). Since prosopagnosia is not a delusion, patients with prosopagnosia do not believe that people around them are faceless. Patients recognize they have a facial recognition problem, and some of them recognize other people by other cues such as clothing, hairstyle, skin color, or voice. In a popular account of misidentification syndrome (Ellis & Young, 1990), prosopagnosia's impaired facial recognition but unimpaired autonomic reaction is said to make a contrast with the unimpaired facial recognition but impaired autonomic reaction of patients with Capgras delusion. In the next section, I will discuss such an account.

7.3. *Existing accounts of misidentification syndrome*

7.3.1. Ellis and Young's account

Ellis and Young (1990) were among the first to give a comprehensive account of misidentification syndrome, and their account is still highly influential. Their account of misidentification syndrome is based on Bauer's (1986) two-pathway account of facial processing, and they explain misidentification syndrome in terms of disruption in one of two face processing streams. According to Bauer, one pathway, whose function is the conscious recognition of familiar faces, is located in the ventral areas and originates from the visual cortex, leading to temporal lobes via the inferior longitudinal fasciculus. The other pathway, called the dorsal pathway, is responsible for autonomic emotional response to faces, and runs from the visual cortex to the inferior parietal lobe, before connecting to the limbic system. Ellis and Young hypothesized the Capgras delusion to be a “mirror image of prosopagnosia.” As I introduced above, some prosopagnosia patients cannot identify persons by face yet they show implicit affective reactions to a familiar face. Therefore, they argue that prosopagnosia patients have a damaged ventral pathway but a preserved dorsal pathway. Patients of Capgras delusion are, on the contrary, believed to suffer a disruption in the dorsal pathway. Disruption in this pathway, as in the case of the Capgras delusion, leads one to not experience familiarity to a close person. This lack of feelings of familiarity seems involved in the belief that the other person is an imposter.

Ellis and Young also gave an account of the Fregoli delusion. In contrast with the Capgras delusion, they think the problem stems from the “cognitive system.” They do not say much about the specification of the cognitive system, but they regard it as comprising of a number of decision mechanisms that makes judgments about the world based on evidence at hand. In cases of Fregoli delusion, the faulty cognitive system overexcites the personal identity nodes for the target of the delusion. Thus the patient experiences persistent thoughts about the target, and their system judges that the target is present. This is consistent with Courbon and Fail’s case in which the patient was an avid fun of theatre. Thus, for Ellis and Young, the Fregoli delusion is more of a top-down effect compared to the Capgras delusion, whose dysfunction is in the autonomic affective systems.

There are, however, a few problems in their explanation of both the Capgras and Fregoli delusions. The problem with their account of the Capgras delusion is that there are people who have defective a autonomic system but do not experience delusions. Some patients with ventromedial region damage report diminished affective feeling to people without experiencing the Capgras delusion. They can still identify people around them in the correct way, despite their diminished affective responses. Thus, a problem in the autonomic system might be necessary for the Capgras delusion to occur, but it cannot be sufficient. This is analogous to illusion: we experience varieties of optical illusions, and sometimes even enjoy them. But we rarely believe the content of illusions when they are inconsistent with our background knowledge about the world or with the testimony from others, as when a stage magician says, “What I’m going to show you is not reality, it is illusion.” However, patients believe in the content of their delusions despite the fact they are often incompatible with other beliefs. Thus, there is something more required to move from “illusion” to “delusion.”

Ellis and Young's account faces the opposite problem with the Fregoli delusion. People can have paranoid thoughts about a particular person, without perceiving them in following them in disguise. To believe that the target person is actually a familiar person in disguise, there has to be something in the perception of the target person, which caused the patient to believe that the target person is a familiar person in disguise. Their account of the Capgras delusion includes an important point here: patients think the familiar person is an imposter because of lack of autonomic response. A natural corresponding suggestion would be that the target person is believed to be a familiar person because of the existence of an (inappropriate) autonomic response. Indeed, this is the suggestion by Ramachandran and Blakeslee (1998). More recent and increasingly accepted accounts tend to include both factors: one sensory, one cognitive. I will discuss one such "two-factor" theory in the next sub-section.

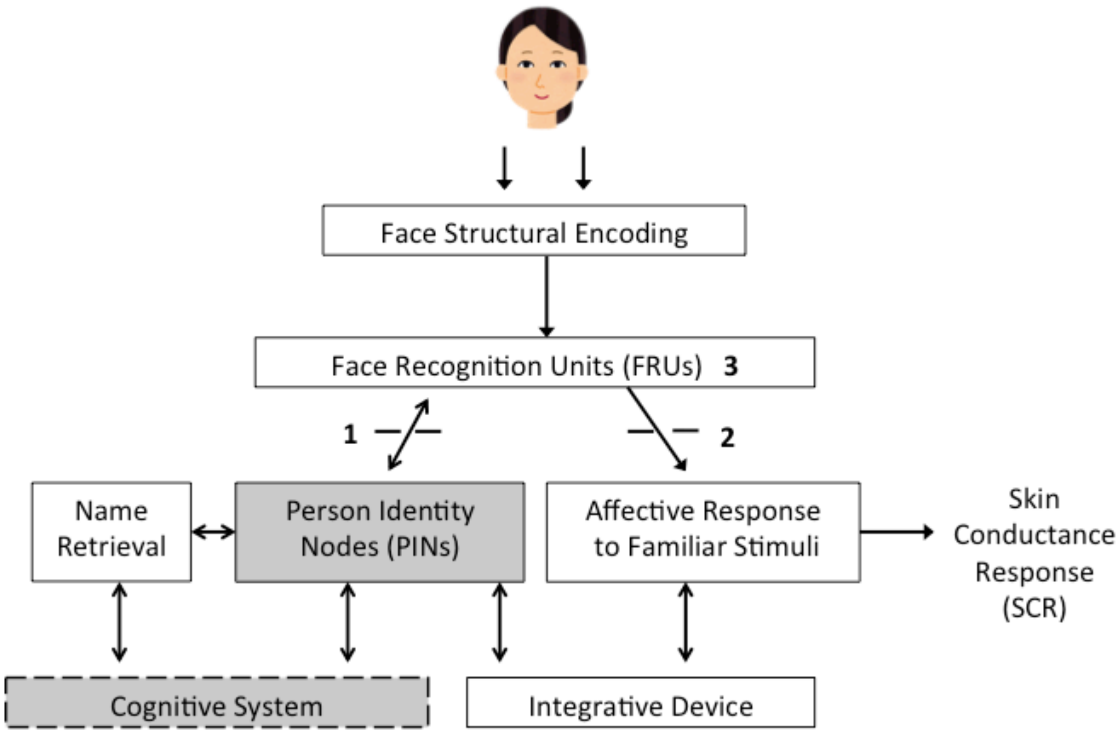


Figure 13 Ellis and Young's scheme

7.3.2. Two factor theory of misidentification delusions

As I concluded in the previous sub-section, modern theories of delusion tend to posit two factors for delusion formation. One factor is sensory, and explains the origin of the content of a delusion. The other factor is related to the incorrigibility of delusions; why delusion is maintained by patients despite counter-evidence. Coltheart, Menzies, and Sutton (2010) give a Bayesian version of two-factor theory. Their *first factor* is also a dysfunction of the autonomic system. In the case of a Capgras delusion, as the representation of the known person has always coexisted with a high autonomic response, "[i]t would be highly improbable for the subject to have the low autonomic response if the person really was his wife, but very probable indeed if the person were a stranger" (Coltheart, Menzies, & Sutton, 2010, p. 277). The low autonomic response gives an explanation for the belief that the person in the patient is not their loved one but a stranger. Recall that posterior probability in Bayes' theorem is proportional to the product of likelihood and prior probability. Even if the prior probability of the hypothesis that the person in front of you is a stranger that looks exactly like your wife is pretty low, the recurrent experiences of absence of high autonomic response slowly and gradually sway the balance. However, as we saw previous sub-section, also argued by Coltheart, Menzies, and Sutton (2010), the delusional beliefs cannot be a good explanation of the sensory experience: no matter how well the hypothesis can explain the *immediate* sensory data, the content of the delusion is not compatible with other beliefs. Patients receive plenty of evidence contradicting their immediate sensory data, including the testimony of surrounding people and advice from doctors. Therefore, Coltheart, Menzies, and Sutton (2010) conclude that we need *the second factor*: failure to take in counter evidence.

The two-factor theory as just explained is compelling, but there still is a lot of space to fill in. Firstly, regarding the first factor, the function of the autonomic system remains unclear. It is argued that sensory aberrance of the patients, namely the breakdown of the co-occurrence of the representation of a familiar person and feelings of familiarity, prompts a subject to stop believing the person in front of them is the who they seem to be. But why this is so? How does the autonomic system pick up a familiar person and represent them accordingly? If the two routes (visual and autonomic) are just redundant systems that do the same job from the same information source, the hypothesis that the person is an imposter is not a good explanation for the experience. The hypothesis that there is something wrong with the brain is at least equally good an explanation for the sensory experience. Thus, the function of the autonomic system also has an influence on how the shape a resultant belief should take if the Bayesian brain hypothesis is correct. Secondly, concerning the specification of the second factor, it is shown that Capgas patients fail to take in counter evidence, but how this actually happens in the brain is not yet developed. Thus, we get two questions below:

1. What is the function of autonomic systems?
2. Why and how does the system fail to take in counter evidence?

Two-factor theories are promising theories of delusion and the Bayesian account proposed by Coltheart, Menzies, and Sutton (2010) is one step toward a predictive processing account. I will now show that the predictive processing perspective provides natural answers to these questions. But before moving to a full predictive processing account, I would like to discuss William Hirstein's mindreading account, because it provides an answer to the first question and leads to more a more comprehensive predictive processing account.

7.4. *Hirstein's mindreading account*

In his recent book, *Brain Fiction* (2005), William Hirstein proposes an interesting account of misidentification syndrome. In contrast with other existing popular accounts, he argues that misidentification syndrome is actually disorder of a certain sort of mindreading capacity. More specifically, he distinguishes two kinds of representation involved in the representation of a person—external and internal representation—and claims that patients with a misidentification syndrome have a problem with internal representation. External representations are about the appearance of an object or person. These are not only about the person's face from the front, but can also be about many various kinds of appearance. As Hirstein (2005, p. 123) states, "[y]ou can recognize their faces at all angles, in poor light, and at some distance. You can recognize their voices no matter what the emotional tonality. You can recognize their walks, their laughs, even their characteristic scents." Thus, external representations may include *counterfactual* representations about the appearance of a given person apart from the representation about their *current* appearance. For example, the content of representation may include that if the person's face were viewed from the side then it would look this way, or that if the person were angry, her face would look that way.

Conversely, internal representations are about typical mental states the person tends to have. Your partner may be a funny and cheerful person; your close friend may be a calm and caring person. When you represent a person, you not only represent their outward appearance, but also their inward personality. Hirstein attributes the phenomenological difference between our experience of a familiar person and of a stranger to differences in internal representations. When you first encounter a person, you can gain information primarily about her outward appearance, but as you interact with her, you get to know more and more and her personality and start representing it. When we are familiar with a person, according to Hirstein, we do acquire feelings of familiarity directly, but rather we acquire richer representations of their inner personality. The crucial proposal of Hirstein's is that the problem experienced by patients with misidentification delusions is in this internal representation. More concretely, patients with the Capgras delusion tie a familiar person with a discordant personality, whereas patients with the Fregoli delusion find the personality of a familiar person in different stranger(s).

One straightforward form of support for this proposal comes from the testimony of Capgras patients.

> One such patient ''became convinced that her husband's personality had changed'' (Frazer and Roberts, 1994). Another claimed there were two doctors looking after him, each with a different personality: ''The first consultant (who he called John Smith) was 'a nice bloke,' whereas the second (a Dr. J. Smith) was someone who was 'distant and aloof''' (Hirstein, 2005, p. 123)

This perspective is interesting as it shows that the imposters appearing in Capgras delusions are not complete duplicates that are qualitatively identical to the original. Rather, they are equipped with different sets of mental states. Hirstein also invokes another description of a patient of self-misidentification to buttresses this hypothesis:

> He developed the delusion that his facial reflection in mirrors and other glass surfaces was no longer his own face but the work of an autonomous evil being. The image of his face was identical to his own face except that the image appeared to be very sinister and aggressive. (Silva et al., 1992, pp.

574–5 as cited in Hirstein, 2005, p.123)

Further direct support for the Fregoli delusion is the containment metaphor that many patients use. According to Hirstein,

> The young woman with Fregoli's syndrome who recognized a friend of hers in every person she saw also "felt that this particular male friend was inside every person. She also felt that her dead grandfather was living inside a nurse.["] (Hirstein, 2005, p. 124)

Hirstein argues this kind of "containment metaphor" is frequently used by patients wih Fregoli delusion. When patients see a stranger, their internal representation of the person is a familiar person's even though the external representation is the stranger's. One interesting consequence of this distinction between external and internal representation, which wasn't particularly focused by Hirstein, is that it is the internal representation that decides who a person *is*. A familiar person with mentality of stranger's is a stranger, whereas a stranger with a familiar person's mentality is a familiar person. This is intuitive, but I will talk more about this point from the predictive processing perspective later.

A summary of the relationship between representation and delusion is provided below (Table 6).

Table 6 Misidentification delusions and external/internal representation

Misidentification delusions	**External representation**	**Internal representation**
Capgras delusion	Familiar person	Stranger's personality
Fregoli delusion	Stranger	Familiar person's personality

What about the neural underpinnings of these kinds of representations? Earlier, I introduced the dual pathway account of facial processing. Hirstein objects to this account by drawing on now widely accepted work by Haxby and colleagues (Haxby, Hoffman, & Gobbini, 2000). According to Haxby and his collaborators, there are two pathways within ventral areas. One pathway including the fusiform face area represents invariant aspects of a face. The other ventral processing pathway involving the cortex around the superior temporal sulcus (STS) produces representations of changeable aspects of faces such as direction of gaze and facial expressions. The representations of the direction of gaze or facial expressions that STS produces itself would be external ones, but the system also plays an important role in forming internal representations. The system that represents changeable aspects is located adjacently to the system that produces representations of objects of interests and expressions of emotion from information about eye gaze and facial expressions. Therefore, Haxby (2000) argues that the ventral pathway involving the fusiform face area is associated with external representation while the ventral pathway involving STS is associated with internal representation. Haxby's (2000) account is largely based on imaging studies, but Hirsterin argues that evidence from monkeys and humans shows dissociations between the pathways. For example, In STS lesion studies in monkey, monkeys exhibit a damaged capacity to understand eye-gaze but a preserved capacity for face identity recognition (Hoffman & Haxby, 2000). Recording studies by Hasselmo et al. (1989) showed that cells in STS were sensitive to facial expression no matter which person the monkey was viewing, and that the cells in the inferior temporal gyrus were sensitive to individuals no matter what facial expressions his or her face showed. A summary of Hirstein's take on Haxby's work is given in (Hirstein, 2005, chapter 5).

Hirstein also maps disruption in these pathways onto associative prosopagnosia and Capgras delusion. Hirstein argues that patients of prosopagnosia have damage in fusiform area but not in superior temporal sulcus whereas Capgras syndrome involves damage in superior temporal sulcus but not in fusiform area. As a support, he refers to his patient with Capgras delusion who has a problem in understanding gaze of a face.

What is the relationship with these two ventral streams and affective autonomic responses? As is introduces above, affective component plays a large role in misidentification syndrome and prosopagnosia. Hirstein argues both streams can produce autonomic reactions since both of them may contact amygdala. Considering the function of the pathways, he argues, "The fusiform gyrus area may be involved in producing an SCR to the sight of a familiar face, whereas the superior temporal sulcus may be involved in producing autonomic responses to emotional facial expressions."

Table 7 Hirstein's take on Haxby et al.'s (2000) two pathways

Two ventral pathways	**Function**	**Representation**	**Dysfunction**
STS-system	Reading changeable aspects of a face Maps facial expressions to emotions	Internal	Capgras/Fregoli
FFA-system	Reading invariant structure	External	Prosopagnosia

One immediate problem arises in Hirstein's explanation of prosopagnosia. Associative prosopagnosia patients exhibit covert autonomic responses when they see a familiar face, including higher SCR, evoked potential, and longer fixation times. If prosopagnosia patients have damage in the pathway involving FFA and if this pathway is responsible for autonomic responses to familiar faces, then proposopagnosia patients should not have those responses. This tension seems to require granting that the pathway involving STS contributes more to autonomic responses to even familiar faces Indeed, Hirstein's analysis of internal and external representation is correct, then patients with Capgras delusion fail to feel familiarity because of a problem with their internal representation. On top of that, Hirstein closely ties internal representation with the pathway involving STS. Thus, it seems he should rather conclude that autonomic responses to familiar faces are also supported by the pathway involving STS.

Hirstein also has a specific account for how we understand the emotions of others. Following work Ralph Adophs (Adolphs, Damasio, Tranel, & Damasio, 1996; Adolphs, Tranel, & Damasio, 2001), Hirstein argues that others' emotions are understood through simulation, which is mainly handed by the inferior parietal cortex. Thus, a failure in reading others' emotions hangs together with a failure of feeling appropriate emotion. This explains why the failure of feelings of familiarity in the Capgras delusion leads to patients ascribing a colder personality to the target their delusion. Hirstein (2000, pp. 126-7) also refers to some cases of coincidental breakdown between feeling one's own emotions and understanding the emotions of others.

> One patient, who alternated between Capgras' and Cotard's delusions, accused the nursing staff of having murdered members of his family (Butler 2000). This man was ''fearful, and perseverated on themes related to death'' (Hirstein, 2005, p. 685).

Hirstein also mentions Wright et al. (1993)'s work describing a patient who exhibited Cotard's syndrome when depressed but exhibited Capgras' syndrome when he experienced persecutory delusions. Patients with Cotard's delusion typically (and contradictorily) claim they are dead. In these cases, covariation between patients' own emotions and the represented emotions of the target of their delusions is observed. Hirstein gives a detailed account of how we come to represent emotions of others. Howevver he is less specific about how we ascribe other kinds of mental states even though he claims that other kinds of mental states are involved both in the phenomenological difference between seeing a familiar face and seeing a stranger, and the phenomenological factors that lead to the Capgras delusion. In one place, he states,

> The difference has to do with the way that you attribute a mind to the person you know. You attribute characteristic emotions and moods, even beliefs. (Hirstein, 2005, p. 123)

And in other place where he is trying to account for a rare case where a patient experienced a Capgras delusion toward their house, he argues,

> The difference is not due to features of the outside of the house, but rather to your beliefs about what is inside your house, and your emotional memories of events that took place there. (Hirstein, 2005, p. 131)

It is unclear how these other kinds of mental states are ascribed in order to yield the supposed pheomoneological difference. Emotional states seem sufficient to account for the phenomenolocal difference between healthy and pathological cases, but in the house case, it is impossible to simulate the house's mind! In accounting for the perceptual mode from the predictive processing perspective, what is important is that behavioral expressions are available to the brain in the short time intervals. Basic emotions are undoubtedly among the ones perceivable, however, beliefs take longer time to be expressed behaviorally. I will discuss how misidentification delusions for nonhuman creatures/objects can be explained in the following sections.

In the remainder of this section, I will evaluate Hirstein's account. Hirstein's account, most charitably understood, can be thought to provide a detailed account of what the autonomic system does. According to his account, disruptions in the autonomic system lead to dysfunction in the pathway involving STS, resulting in an unmatching internal representation. This provides a potential answer to the first question introduced before. However, this account has several problems.

Firstly, his account needs more direct empirical support. Hirstein refers to some cases of Capgras delusion including, his own patient, but there is insufficient evidence for the Fregoli delusion and other misidentification delusions. The evidence wanted for the Fregoli delusion would be similar to that provided for the Capgras delusion: covariation between a patient's own feelings and the represented mental states of their targets.

Putting this point aside, the most problematic issues surround the very distinction between internal and the external representations. The name "internal representation" and the containment metaphor used in Hirstein's account of the Fregoli delusion give the impression that the internal representation is not tied to observable expressions, but this is not the case. Indeed, it is unclear how we can even have internal representations of others if mentality cannot be expressed behaviorally. Hirstein himself argues that emotional states (if not all mental states) can be read from the changeable aspects of someone's face. At the same time, some forms of the content of external representation can also be considered "hidden". In Chapter 2, I discussed representation as of a whole object. I argued that representation as of a whole object cannot be determined in a purely bottom-up way; it has to be inferred, as many different objects can give rise to the same sensory stimulation. By the same token, many different faces or non-face objects could give rise to the current sensory stimulation one is getting from a face. Thus even some contents of external representation are hidden from our current sensory stimulation. Consequently, the distinction between hidden and exposed does not correctly map onto the distinction between internal and external.

I suggest that we should instead talk of the distinction between dynamic and static content. This distinction is suggested because it directly coincides with the function of the STS pathway, and moreover, mental states are considered tightly related to actions, which are essentially dynamic. Mental states themselves are static contents, they can be expressed or inferred by dynamic contents. We already discussed the case of basic emotions that are directly readable from facial expressions, but even fully-fledged propositional attitudes have a tight, if complicated and indirect, relationship with actions. For example, if I see you, on a hot day, reaching toward a glass of water on a table, I would ascribe not only the intention to drink water, but also the belief that there is a glass of water on the table. The inference is not deductive, but rather an inference to the best explanation.

Similarly, representations of a whole face or a part of a face themselves are about static and invariant aspects faces. However, they are also explained by a certain kind of dynamic content, such as the appearance of a face seen from many different angles or lighting conditions. Indeed, static contents are essential to representing dynamic aspects. When representing lips that are moving, it also has to be represented that the same lips exist throughout the changes! Using the distinction between dynamic and static contents rather than distinguishing between internal and external representation has at least two virtues. For one thing, this distinction does not need the containment/hiddenness factor. Dynamic contents just take time to become apparent; they might not be grasped at an instant. For another, it is apparent that the distinction admits of degrees. The dynamicity of contents may be determined by the timescale at which the dynamics are defined. Recognizing facial emotional expression may take only a few seconds, whereas recognizing my intention to cross a pedestrian bridge may take more time. This distinction between dynamic and static contents is easily transposed to *Hierarchy*, as I will discuss this later in this chapter.

A second related point on Hirstein's distinction is that it can be questioned whether the problem experienced by patients with misrepresentation delusion is really in internal representation or mindreading. In a recent paper, Langdon et al. (2014) point out the apparently intact mindreading abilities of patients with the Fregoli delusion, as they make inferences about the motives of the targets of their delusion. Langdon and colleagues therefore demand further supporting evidence for the mindreading account.
In one sense, the label of "mindreading theory" might be misleading. Where Hirstein locates the problem is not in the mindreading system per se, but it is in a ventral

pathway that is responsible for the changeable aspects of a face and provides *input* to the mindreading system. Therefore, it is no mystery that patients of Fregoli delusions try to make inferences about their targets' mental states. Therefore, even though the most bizarre (hence, outstanding) abnormality is in the content of internal representation (the represented personality of the target person), the main disruption may rather be in the external representation (the changeable aspects of the face). This provides us with another reason to take up dynamic/static distinction, as the problem in misidentification syndrome seems to be in understanding dynamic sensory contents.

The last point I would like to consider here comes from Pacherie (2008). Her consideration further pushes toward the distinction between dynamic/static contents. She questions why the erroneous ascription of emotional states leads patients to form the belief that the target person is an imposter rather than the belief that she is only in a bad mood for some reason unknown to the patient. Pacherie gives two kinds of reply to this objection. The first kind focuses on the fact that the changed personality is not temporary, it persists through time, and this temporal accumulation might lead one to succumb to delusional beliefs. This may be true, but it is unclear whether or not such patients have embraced their delusional beliefs "straight away." Her second answer looks more interesting to me:

> [T]he second line of answer is that impairment to the lateral temporal pathway would disrupt not only the correct reading of expressions of emotions but also the identification of *the dynamic signature of the face of the person* [emphasis added]. Someone with such an impairment would not only mistake his father's expression of concern for one of anger, but would also see this expression of anger as different in its dynamics from his father's ordinary way of facially expressing anger [emphasis added]. (Pacherie, 2008, p. 113)

Here Pacherie brings up something closely related to but beyond mental states. She talks about the particular way an emotion is expressed, or an action is carried out, by a particular person. For example, one person might always smile by raising one end of their lips, another might always show her teeth when smiling. There are subtle differences from person to person in what smiling raising looks like. The dynamics or trajectory of executing one type of action would probably be unique for a particular person. Putting Pachrie's worry differently, perceivable types of emotion alone probably fall short of specifying a person. Appearance alone also falls short of specifying a particular person, and likely appearance plus perceivable mental states also underdetermines particular persons. This also speaks to the dynamic/static distinction. The problems in misidentification syndrome may go beyond the realm of mental states, and alsoinclude more dynamic aspects.

To sum up, Hirstein's account provides a unique alternative perspective on the role of autonomic function. However, the usefulness of his distinction between internal and external representation that his account rests on is questionable. I have suggested that the problem experienced by patients with misidentification syndromes should instead be framed in terms of dynamic contents. In the next section, I discuss one such new theory of personal identification and show that it is compatible with the predictive processing framework.

7.5. *Dynamic identity signature and personal identification*

The previous section called for attention to the more dynamic aspects of person identification. I suggested the way a person executes an action forms a "signature" or "fingerprint" of that person; it is a reliable and quick source of identification while not being something consciously considered. Human beings (and certain other animals) are sophisticated mind-readers, and healthy adults can distinguish differences in mental states from subtle differences in physical appearance and expression. For example, healthy adults can normally distinguish real smiles from smiles consciously made, or a person being entertained from a person pretending to be entertained. It is not so surprising that humans can pick up someone's personal signature from the way she behaves. It has also been demonstrated that humans can perceive biological motion from the motion of arrangements of dots that preserve some features of biological motion. In Point Light Display (PLD) studies, human biological motion is recorded by lights attached to the joints (Johansson, 1973). These studies have shown that the information obtained by the perception of the movement of the dots is sufficient to perceive the human motion. It is interesting that humans can identify a person just through the motion of dots that preserve features of that person's particular ways of executing actions.

Despite this apparent importance of the dynamic aspects of human behavior, they mostly remain unexplored in cognitive science. This is because of three biases in the area: Static images are mainly used in the study of face perception, dynamic bodily motion is mainly studied for action perception and social communication, and the human voice is usually studied for voice perception. However, a recent body of research shows the importance of dynamic aspects of human behavior in human identification.[53]

Yovel and O'Toole (2016) reviewed the literature on the contribution of three kinds of dynamic information (facial motion, bodily motion, and voice) to person identification. They argue that all the kinds of dynamic information mentioned above contribute to person identification, but the size of each's contribution depends on the context. For example, bodily motion plays a leading role when viewing conditions are suboptimal, whereas facial motion is predominant when the person is familiar but image quality is not high. However, perception of the whole body in motion gives the best result in identification. The contribution of motion to person identification also depends of the kinds of motion. Based on studies with PLDs, it is greater for certain types of action, such as dancing and boxing, than for other types of movement, such as walking. Nonvisual information, such as voice, is also said to play an important role in person identification. For example, in one study, participants could match muted video of an unfamiliar person speaking with that person's voice, even when the voice and the face said different sentences (Rosenblum, Smith, Nichols, Hale, & Lee, 2006). Given the intrinsically dynamic nature of voice, this suggests a multisensory integrative system for person identification (Yovel & O'Toole, 2016).

Furthermore, the level of familiarity is an important variable for person identification. For example, dynamic identity signatures are more likely to facilitate the recognition of familiar rather than unfamiliar people, because we have prior experience of familiar people's idiosyncratic movements and gestures. That the integration between voice and face is stronger for familiar persons is suggested by psychophysical studies as well as an electrophysiological study (Yovel & O'Toole, 2016).

Interestingly, the core of the neural mechanism underlying these dynamic aspects

[53] Indeed, Haxby et al. (2000) postulated that the pathway that underlies the representation of the invariant aspects of a face is responsible for facial identity, but that the pathway that underlies changeable aspects is not. However, in their more recent work, Haxby et al. (2011) admit the role that changeable aspects play in identification.

is suspected to be the STS, the area also discussed in Hirstein's account. As I have already discussed, the STS is thought to be involved in the representation of changeable aspects of voice, and it is also thought to be involved in voice recognition. Thus, Yovel and O'Toole suggest that the system that involves STS is a multi-sensory integrative mechanism.

In addition, this emphasis on sensory signature is also supported by evidence from developmental psychology. Andrew Meltzoff (2005, 2007) has proposed a "like-me" hypothesis. He argues that imitation plays a fundamental role in developing social cognition and that infants' regard for other persons as "like-me" is the basis for this development. He also argues that infants learn to identify particular persons by imitation. When infants are not sure about the personal identity of persons they encounter, infants imitate how the persons behaved at the first encounter (Meltzoff, 2007). This hypothesis is further supported by a new machine learning study (Boucenna, Cohen, Meltzoff, Gaussier, & Chetouani, 2016). In Boucenna and colleagues' experiments, robots and humans imitated each other, and through this, the robots built up behavioral models of individual people. After this training, the robots could correctly identity each person at a high rate. In these two experiments, the idiosyncrasies used to identify each person could not be their mental states; the main action of the experiments was imitation and no behavioral expression of mental states was possible. Thus, this evidence suggests that humans come to track personal identity through tracking individuals' subtle and unique ways of executing an action.

Identification based on dynamic signatures is analogous to expert perception. This account shows that we automatically use various cues to make more accurate identification judgments. We are also often unaware of the causes of these identification judgments. Yovel and O'Toole (2016) note that participants are typically unaware that they rely on the body when making an identification judgment. This fits with the characterization of intuitive knowledge of experts as automatic. The intuitive knowledge of experts is also said to be domain specific; we are experts at identifying persons, but most people are not experts at identifying individuals of other species.

This section summarized the recent literature focusing on the impact dynamic information has on person identification. The Yovel and O'Toole's theory claims that dynamic information plays an important role in person identification and this viewpoint is exactly the one wanted in the previous section. Given this point, however, what can we say about the problem that the patients with misidentification syndrome have and the role that the autonomic system plays? In the previous section, I suggested that the problem experienced by patients with misidentification syndromes is in the mapping from dynamic contents to static contents. However, if both the invariant aspects of a face and mental states are static aspects of the mind, and if patients do not have a problem grasping the invariant aspects of a face, where exactly does the problem of patients with misidentification syndrome lie?

There seems to be an important difference between grasping the invariant aspects of a face and identifying a person. Representing personal identity or mental states from motion or behavior essentially involves dynamic patterns unfolding over time, whereas representing the invariant aspects of a face does not. Picking up anger in the face requires picking up changes in the face; picking up personal identity from dynamic signatures requires picking up subtle idiosyncratic patterns that unfold over time. Conversely, representing the invariant aspects of a face relies on is picking up parts that do not change over time. In other words, representing personal identity or mental states is a more difficult inference problem because of the time-extended nature and way in which cues change over time. This also provides an answer to the first question. The autonomic system contributes to the ventral stream involving the STS and we can speculate that patients with misidenfication syndromes have problems in this route, which underlies the mapping from dynamic information to mental states or personal identity. It may be that the error is in representing motion or behavior and that this subsequently leads to the misrepresentation of mental states and person identity. However, we are still wanting an answer for the second question: the failure to take in counter evidence. To answer this question, we will look into the person model suggested by Albert Newen, along with the two-system view, and integrate them into the predictive processing perspective.

7.6. *Putting all things together: a predictive processing account of person identification and misidentification delusions*

In a recent paper, Albert Newen offers a new theory of understanding others. His theory is the *person model theory* (PMT) and it posits two kinds of person models used in understanding others (Newen, 2015). The first model is the person schemata. A Person schema is an automatic system whose processes usually remain unconscious.

Newen states,

> A person schema is an intuitively formed, implicit model of a person; it is a memorized unity of characteristic features of a person including facial features and expression, voice, moving pattern, body posture, gestures, and other perceivable features of a person. (Newen, 2015, p. 13)

Thus, person schemata include the dynamic signatures just discussed, though this is not focused on. Newen argues that person schemata are used for the quick assessment of others in aspects important for survival such as familiarity, safety, or attractiveness.

The second system is the person image. The person image is a model of a person whose contents tend to be available to consciousness and are linguistically expressible. This is developed at a later stage based on the person schemata. The person image can also be developed through narrative. One of the main reasons why Newen thinks there are two systems stems from misidentification syndromes. He sees misidentification syndromes as consisting in the malfunction of person schemata, because of existence of largely correct conscious knowledge about the person and unconscious nature of affective processes, which are the main culprit of misidentification syndromes. He argues that PMT is compatible with two-factor theory (discussed in Section 7.3.2.), as two factors seem to map on to the person schema and the person image. The second factor has to do with why the person image trusts the output of person schemata in the face of a large body of evidence to the contrary.

In this way, the PMT gives a comprehensive account of understanding persons even though it is not focused on person identification. However, PMT does not itself state three things in detail:

1. How are person schemata organized?
2. How do person schemata and person images relate to each other?
3. How does the dysfunction of a person image lead to a misidentification syndrome?

Newen's PMT seems compatible with the work that claims we have two modes of personal identification processes (Bullot, 2014). Personal identification can be required to operate in one of two contexts. In the first sort of occasion, personal identification is straightforward and seems "perceptual." When I go to a philosophy seminar and see familiar faces, I straightforwardly and without any conscious effort identify the people surrounding me. In other sort of context, however, personal identification is more laborious. A police investigation to identify a murderer could be an example. A more daily example is the following: when I wonder about the identity of the person whose back I saw on campus yesterday, thirty meters away from me under dim light, I cannot see who she is immediately, I have to think about my friends and their schedules consider the possibility that it could be one of my friends.

Also, the PMT is similar to two systems view of the mind proposed by Stanovich and West (2000) and developed by Kahneman (2011). According to this view, our cognitive system is made up of two systems that have different characteristics. System I is more automatic, quicker, and unreflective, and its processes tend to be unconscious. Conversely, System II is more and elaborate, slower, and reflective, and its processes may be conscious. Note that this characterization regarding consciousness only pertains to processes not to the products of the systems. Thus, the end product of the System I may as well be conscious. For example, we usually don't have access to the processes within our visual systems but we do have a conscious percept. System I provides a domain specific, "quick and dirty" mode of reasoning, whereas System II provides a slower but more domain-general way. Clark provides a recent treatment of two systems view from the predictive processing perspective, andI will answer the last two questions using Clark's work later in the chapter.

Before that, we can work on the first question using *Hierarchy* from predictive processing. *Hierarchy* gives us a tool to organize person schemata. Earlier, I proposed that dynamic/static contents can be transposed onto the length of time over which a model makes predictions in the predictive processing scheme. We can organize the different kinds of representation we have discussed so far in a hierarchy as in Figure 14 below.

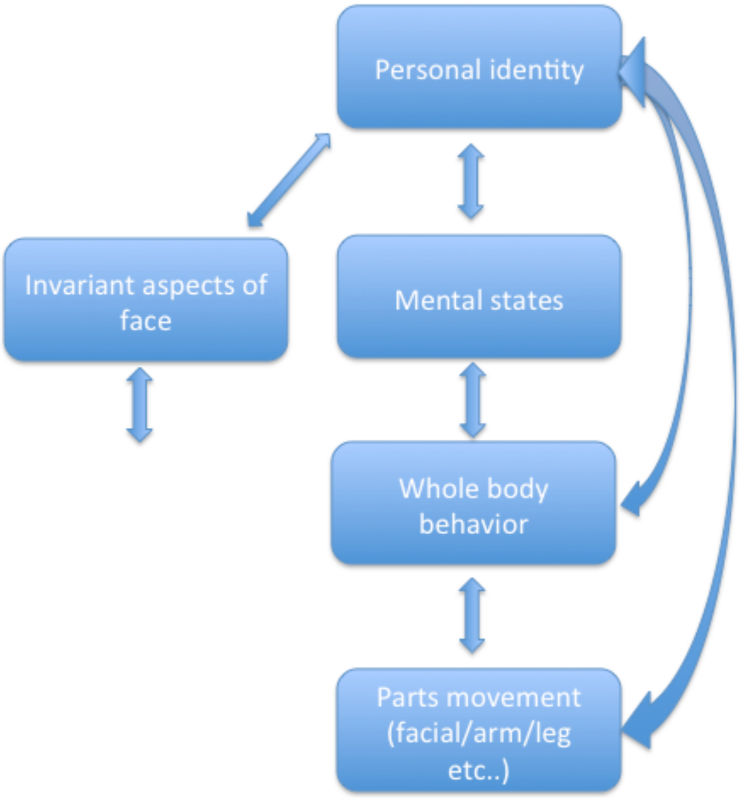


Figure 14 A model of the hierarchy of a person

The generative model of personal identity would be situated at a higher level given the timescale it concerns and its level of abstractness. It makes a prediction about the perspective invariant aspects of a face or other parts of a person, the typical mental states that person would have, and the mental states that person has in that situation. In turn, the representation of the mental states makes a prediction about what kind of behavior the person is performing.

Usually, the predictions made by personal identity representation or by mental states are quite imprecise ones, as it is usually very difficult to predict mental states from your personal identity and background knowledge about a situation. (For example, you might be thinking about dinner plans when you are having a chat with someone but it is hard to guess about the other party). However, some kinds of mental states in some situations enable quite precise predictions. For example, it has been argued that basic emotions and motor intentions are perceivable. In predictive processing terms, this means those kinds of mental states make precise predictions on a short timescale. In the same vein, the most important point in this hierarchy is that the representation of person identification also makes a precise prediction about behavior or the motion of parts. This is one of the central claims made in this chapter derived from the dynamic signature account of personal identification.

The hierarchical model presented here is similar to the predictive processing model in Chapter 4, and to reverse hierarchical theory (Ahissar & Hochstein, 2004). In the account and the theory, the high-level gist representation is formed earlier than the low-level visual representation. The high-level gist representation is formed based on crude visual information, and based on the high-level gist representation, more detailed information is explored. In a similar way, recognition of identity is made based on sensory cues. Although this has not yet been empirically explored, it is possible that recognition of identity comes earlier than predictions about other features.

Moreover, this hierarchical view is related to Noë's sensory-motor theory of perception discussed in Chapter 2. In Noë's account, perceiving objecthood (perceiving the appleness of an apple for example) is to know how its appearance changes depending on one's action. This knowledge is not propositional knowledge, but rather a kind of know-how, which may be difficult to put into words.

Identifying a particular person can be accounted for similarly. Identifying a particular person consists in having knowledge about how the person appears depending on your actions and theirs. This knowledge would include complex causal patterns observed through social interaction between the perceiver and the person. This is sometimes conceived in terms of long-term actions such as approaching or grasping, but it can also conceived in terms of short-time actions such as saccades. The latter type of sensory-motor contingency can correspond to predictions about the sensory signature. The prediction has to be precise enough to pin down the person, but the precision of the prediction would probably depend on how familiar one is with the person being identified. If the person is a business-related person who you meet only once in a week, for example, your model may not be able to produce very precise predictions. It would include how she usually dresses and how her face looks, but might not include specific information about how she smiles. However, if she someone close to you who you meet every day and spend a lot of time with, the predictions made by the model would be highly precise. They would include the characteristic ways she behaves. The high-level model of the person therefore modulates precision expectation at the low-level so as to predict a sensory signature.

This point is important also because the Capgras delusion only seems to happen to familiar persons, animals, things or places. The model of a particular familiar person can be quite rich; it makes predictions over many different timescales, and so there is more leeway for the model to go wrong. The model of a particular familiar person is a high-level model. It makes predictions about lower level models, about things such as the person's outward appearance and personality. So it makes predictions about how the person will appear from a certain viewpoint and in certain environmental conditions, and also makes predictions about types of mental states she is likely to have at a given moment. This in turn leads prediction about how she will behave. This prediction of behavior from models of mental states is not a precise one as the relationship between mental states and behavior is not straightforward and mental states do not pin-down any specific way an action or behavior will be performed. Being angry does not lead you to behave in any particular way; sometimes you might express anger in an obvious fashion, but other times you have to suppress it. Furthermore, there are many ways to make a grimace and the prediction only from the model of anger is not sensitive to that variety. On the other hand, the model of a particular person makes precise predictions on models at lower levels about the motion of that particular person by specifying how an action is performed by that person.

Earlier, I connected this account of personal identification with object perception in my dynamic sensory account and expert perception, but there is a noteworthy difference between expert perception and personal identification. Targets of expert perception are usually multiple, be it detection of chicken sex or of elm trees. There are many female or male chickens, and countless elm trees. However, identification of a person essentially involves only one person. Identifying Ryoji Sato is not picking out persons who share that name. It is to pick out one particular individual. This difference needs to be addressed.

I do not think the difference between usual cases of expert perception and personal identification is crucial. The difference seems to be just a matter of degree. It is a constraint on the model that it has to be precise enough to pick out only one person in the vicinity of the perceiver. The system has to be able to take in the differences between two persons of roughly the same size with the same hair color, eye color and so on, to be able to represent each particular person. Almost all of us can be pretty good at this, to the level that family and close can distinguish between identical twins spontaneously. However, the system is not infallible. As I discussed in Chapter 4, if an object is swapped with a sufficiently qualitatively similar object, there is no way for this system to register its numerical difference. One strength of the theory is that in cases of Capgras delusion, it can accommodate the fact that there is no qualitative difference. The theory predicts that there also is a difference in how sensory information about the person is processed.

Even though I argued that the difference between expert perception and person identification is just a matter of degree, this difference is nevertheless related to the character of misidentification delusions. Person identification is a highly private phenomenon, who can identify whom differs from person to person. More likely than not, you are one of the few experts in the world in quickly identifying your partner, parents or people in a similar position. Indeed, studies on the impact of dynamic information on personal identification suggests that familiarity is an important variable. I already mentioned that sensory signatures are only available for familiar persons, and that voice/face integration is also more available for them.

This point is related to the second factor: failure to take in counter evidence, which I will discuss shortly. Earlier, I argued that this problem is in the sensory domain and it leads to errors in mapping from behavior to mental states and personal identity. The predictive processing account reveals the first factor of misidentification as consisting in some problem at a lower level model in the hierarchy. In this way, the person schemata and dynamic signature account can successfully be accommodated in terms of predictive processing.

This picture also coincides with an existing predictive processing account of delusions. Hohwy (2013) gives a general account of delusion from a predictive processing perspective, focusing more on explanation of the second factor. Hohwy also explains the incorrigibility of delusion in terms of reality testing. Deluded patients keep holding delusional beliefs despite the existence of testimony of other people and despite general background knowledge incompatible with their delusional beliefs. Reality testing is a procedure used to check prior beliefs against other kinds of evidence. It is usually undertaken when “perceptual content is unexpected, or unexpectedly uncertain; sometimes we subsequently learn from others or from later evidence, that a perceptual inference was misguided” (Hohwy, 2013, p. 148). Thus, the incorrigible constant upholding of delusional beliefs can be explained by a failure of reality testing. To see why it fails, it is first necessary to see how it works. Consider Figure 15 below. Upon seeing the picture for the first time, you might regard it as consisting only of black and white patterns. However, upon learning that the picture contains two cats, you would spot hidden cats after a few moments; your percept will suddenly change. Now, you cannot help but see a Dalmatian in the bottom right corner.

Figure 15 What is it in here?

Reality testing works in this case because the hypothesis that there is a Dalmatian translates into a lower-level model of what a Dalmatian looks like, and this leads to an explanation of the low-level prediction error. However, if high-level hypothesis fails to be translated into prediction at low-level, the model that is actually more accurate fails to win. In the case of delusion, the most accurate hypothesis that the cause of the delusion is some dysfunction in the brain. This is attested to the patient's doctor and is compatible with testimony of other surrounding people. However, this hypothesis yields to the delusory hypothesis that, say, the CIA is chasing and trying to kill her, because the former hypothesis does not explain lower-level, more sensory prediction errors, whereas the delusional hypothesis succeeds at this task.

How does reality testing work given the dynamic signature account? In the case of Capgras and Fregoli delusions, reality testing fails because background knowledge and the testimony of others fails to be translated into low-level prediction errors. But this sensory dynamics account gives an interesting twist in the account. According to the account presented here, the high-level model of personal identity does translate into the low-level prediction error and this translation is only available in the case of familiar persons. In the case of the Capgras delusion, it is arguable that because of lower-level sensory errors, patients fail to see the sensory signature of a particular person, and thus they cannot identify that person as familiar. In a similar way, in the case of the Fregoli delusion, sensory low-level errors somehow lead to recognition of sensory signature a person who is not present. The Fregoli case may be trickier to incorporate into this model, as a more top-down account is available. It might instead be due to abnormally higher precision expectations for a familiar person. This at the same time leads to abnormally low precision expectations for low-level prediction errors, leading the patient to disregard nonmatching sensorium and thereby hallucinating the sensory trajectory of the object of their delusion. This more top-down account is compatible with the accounts of Young and Ellis (1990), and Langdon et al. (2014).

I argued above that in the case of the identification of familiar persons, you are almost the only expert in the world. This is another kind of evidence insulation; good reliable sensory evidence to identify the particular people you know well is only available to a limited number of people. This adds another explanation for why misidentification delusions are incorrigible; hypotheses provided by other less knowledgeable people are not estimated to be very precise.

Hirstein tries, using his mindreading theory perspective, to explain why misidentification delusions sometimes occur to familiar nonhuman animals and even to familiar inanimate objects such as a patient's house. This theory can readily account for cases concerning nonhuman animals, but it remains unclear how his mindreading theory can explain the house case. To meet this challenge, he argues that we also have internal representations for a house. He states,

> I suggest that we form internal representations for more than people. […] Externally, your house looks almost exactly like the others. But for you, looking at your house is a completely different experience from looking at the ones nextdoor. The difference is not due to features of the outside of the house, but rather to your beliefs about what is inside your house, and your

emotional memories of events that took place there. (Hirstein, 2005, p.131)

However, to support this position, he would have to tell different stories regarding how this can take place, because such cases do not seem to involve any of the sort of understanding of motion or the ascription of mental states that his theory is based on. He tries to support his relaxation of the conditions of internal representation by appealing to the empirical finding that the cortex close to the fusiform gyrus responds to nonhuman faces and faceless animals (Hirstein, 2005, p. 131). However, both the case of the Capgras delusion about the house and this evidence seem better explained in terms of the predictive processing account. It is important to note that even though the Capgras delusion can, in rare cases, target nonhuman objects, such cases are confined to deeply familiar objects. Thus, it may be argued that the Capgras delusion can target a particular person's house because the house is an object that the patient has expertise in identifying. This can be explained in terms of the expert perception view of the function of FFA (Gauthier, Skudlarski, Gore, & Anderson, 2000).

There are two opposing views about the function of FFA. A traditional view is that the area specializes in face-related processing, but more recent evidence suggests that the area deals with those objects that a person has expertise in recognizing. If the two pathways in Haxby's theory can be translated to expert perception terms, the pathway involving FFA may be responsible for the more static aspects of expert perception whereas the pathway involving STS may be responsible for expert perception of more dynamic features.

This predictive processing account of misidentification syndrome is still wanting more direct empirical support, but it is an interesting alternative which makes testable predictions. It is also a natural progression from the basic idea of predictive processing and the two-system view. According to the base idea of predictive processing, what the brain does is to capture hidden causes in the world. The particularity of a person is hidden in sensorium and the brain has to pick causal regularities hidden in different timescales. We have now answered question 1 above about the function of autonomic function and the second factor.

We still have two questions left to answer:

2. How do person schemata and person images relate to each other?

3. How does the dysfunction of a person image lead to a misidentification syndrome?

I will now answer these questions from predictive processing perspective, using Andy Clark's recent work on the two system view. From the perspective of PMT or the two systems view, the problem can be conceived as a switching problem from person schemata (System I) to person image (System II.). Normally, we have no problem in switching from person image to the reflection mode (from System I processes to System II processes). If the situation is an "easy" one where the situation and the target person is familiar, the personal image is sufficient for the identification task. However, if the situation is "difficult," for example, if the environmental conditions are non-ideal or if you don't have much experience with the target person, you would have to engage in thinking or exploration to identify the person. Usually, even this switching is effortless; we don't have to decide when to think more consciously. Patients of misidentification delusions seem experience failures when switching between these two mechanisms. Patients with misidentification syndrome have delusions because they accept the output of faulty perceptual schemata, when they should turn down that consciously by consulting person image. Even though inner processes of person schemata are almost always unconscious, their output of person schemata may be conscious. For example, the conscious perception that the person is an imposter or the person is a familiar person in disguise. However, as I argued before, they should not highly evaluate the content of this representation. How can person schemata, person image, and the switching problem be explained in terms of predictive processing?

Andy Clark (2015) discusses similar points from the predictive processing point of view in his recent work. Clark explains how the brain chooses the right learning strategy incorporating contextual information. He starts from "model-based" and "model-free" approaches of learning and adopts a recent integrative approach to them. Model-based approaches include a rich model of the task domain so as to make it possible to estimate the value of possible actions. Model-free approaches are devoid of such models, instead working more directly on perceptual data, and "[s]uch approaches implement "policies" that typically exploit simple cues and regularities while nonetheless delivering fluent, often rapid, response." (Clark, 2015., p. 13). Since the model-free approaches do not incorporate explicit models and instead make use of simple cues and regularities (quick and dirty methods), such approaches would sometimes go awry.

The model-based approach corresponds to System II (the person image) and the model-free approach corresponds to System I (the person schemata). However, Clark argues the dichotomy no longer holds and model-based and model-free approaches can be captured under a "more integrated computational architecture" (Clark 2015). Under this integrative view, the brain areas associated with both model-based approaches and model-free approaches interact with each other. Top-down information uses contextual information to control how these areas are combined. This integrative view has an obvious echo with predictive processing. Clark writes,

> Within the PP framework, this would follow from the embedding of shallow "model-free" responses within a deeper hierarchical generative model. By thus combining the two modes within an overarching model-based economy, inferential machinery can, by and large, identify the appropriate contexts in which to deploy the model-free ("habitual") schemes. (Clark, 2015, p. 14)

This integration of model free responses and a model based approach is achieved through manipulation of *Precision Expectation*. Precision expectation plays a role in determining the extent to which prediction error is taken into account, but Clark highlights another role.

> [A]nother important role is the implementation of fluid and flexible forms of large-scale "gating" among neural populations. This works because very low-precision prediction errors will have little or no influence upon ongoing processing, and will fail to recruit or nuance higher-level representations. […] When combined with the complex, cascading forms of influence made available by the apparatus of top-down prediction, the result is an inner processing economy that is […] "maximally context-sensitive". (Clark, 2015, p. 13)

That is, the higher-level model controls how processing at the lower-level goes, by modulating precision expectation to reflect the context and choose the most suitable strategy. The predictive brain, via optimal precision expectation and hierarchical structure, uses context to decide when to use the perceptual schemata and when to use the person image. The "embedding" of the model-free response in the model-based approach is important to understanding the relationship between the person schema and the person image. Although Newen does not talk explicitly use a hierarchical relationship, put in hierarchical terms, we can say person schemata can be positioned at lower levels than person images. Person images involve linguistic representations, which are usually located at higher levels, whereas person schemata involve more perceptual processes, which are usually located at lower levels. It is important point to note that at some levels, the person image and person schema overlap and can involve conflicting contents located at the same level. This is exactly what happens when a patient fails to take in counter evidence. In those cases, the output of the person schema is incompatible with other beliefs. Here we can invoke reality testing again to explain why the output of the person schema becomes the winning hypothesis: the output hypothesis can explain the sensory aberrance whereas other hypotheses cannot be understood in terms of low-level prediction errors. To summarize, I suggest that from the predictive processing perspective, the person schema is located lower than the person image. Further, the upholding of the output of person schemata can be explained in terms of reality testing.

7.7. *Answering questions*

I will now summarize answers to the four questions from Chapter 1 to conclude this chapter.

Answering the inverse problem

Personal identity is the deepest perceived aspect of the world among the five Ps, and it was initially unclear how it could be manifested in sensorium. However, it turns out that we can rely on multiple sources of information—what someone's face looks like, how they make facial expressions, the specific ways they behave or move, and so on. Some of these factors involve dynamics: action related information that cannot be manifested in sensorium at an instant. Together with our usual reliance on priors, the inverse problem of representing personal identity is solved by relying on multiple sources of information involving differing timescales.

Answering the description and perceptuality problems

One difficulty in describing the phenomenology of the experience of persons is to decide its perceptuality. I therefore answer two problems together here. Sometimes the phenomenology of persons seems more perceptual, sometimes it seems more cognitive. It might be argued that personal identity recognition cannot be perceptual given the depth of its causal hierarchy. However, I identified two modes of recognition of person identity, one relies on person images and the other relies on person schemata, and explained both modes from a predictive processing perspective. The predictive processing account supports the perceptuality of recognition involving person image and the non-perceptuality of recognition involving person schemata. The person image is situated at lower levels in the hierarchy and is called upon to explain sensory stimulation. This fits with the definition of perception given in Chapter 1. In contrast, recognition by person schemata is performed when the context is not good enough for representation. In this case, precision expectation at sensory stimulation is low and the processes relied on are more on top-down.

Answering the reason problem

Why do we experience personal identity? We started from the biological importance of identifying individuals: it is advantages for an organism to be able to represent the identity of others. However, doing so seems a difficult job for the brain. The account provided in this chapter demonstrated how the brain relies on multiple sources to solve this problem. As we have seen, it has been frequently argued that affective processes are involved in person recognition, but it has not previously been made clear what kind of information, apart from facial information, is used for identification. My account sheds light on the issue and shows the importance of dynamic information.

The account provided here also explains why the recognition system sometimes fails. Personal identity is a deep aspect of the world. Thus if the balance between reliance on sensory information and cognition is not well made, the systems collapses. Misidenfication syndromes take place when the system relies on sensory information when it is supposed not to.

7.8. *Conclusion*

In this chapter, we discussed personal identification and disorders relating to it. Seen from the predictive processing perspective, personal identification involves tracking the dynamic aspects of a person. When we can directly see the person, we are able to track idiosyncratic sensory dynamics of that person. This dynamic sensory aspect of person identification has only recently been explored and it has not been applied to misidentification syndrome. This has provided an explanation of misidentification syndrome incorporating this sensory aspect with other predictive processing work on delusion.

8. *General Conclusion*

This thesis has focused on five Ps in conscious perception: *Presence*, *Poverty*, *Present*, *Particulars*, and *Persons*. Despite its initial appearance as diverse phenomena, commonalities amongst the Ps have transpired when seen from the predictive processing perspective. Here, I briefly recap the findings from each chapter, to bring these commonalities into focus.

In the Chapter on *Presence* (Chapter 3), I proposed the tripartite representation account of perceptual presence: representation of seen parts, representation of unseen parts, and representation as of the whole object. The tripartite representation was explained in terms of hierarchical Bayesian inference.

In the Chapter on *Poverty* (Chapter 4), I discussed the problem of the richness of visual consciousness. Through careful analysis of the phenomenon, it turns out that it is another rendition of the inverse problem: how can we represent the world given limited sensory evidence? It turned out predictive processing framework is equipped with the tools to solve this problem thanks to its prediction error minimization system.

In the Chapter on *Present* (Chapter 5), we discussed problems relating to the specious present and I discussed multiple "now" hypotheses to explain the specious present. I proposed that the specious present actually consists of representations at different timescales. The analysis itself naturally called for the hierarchy in predictive processing. Importantly, the predictive processing account for Present is similar in structure to the accounts for Presence and Poverty, despite some superficial differences. This similarity allows a new and intriguing empirical prediction, namely of the existence of a temporal version of change blindness within the specious present, which should be explored in future research.

In the Chapter on *Particularity* (Chapter 6), I dealt with representation of particular objects. The insensitivity of mechanisms for representation of particular objects from conscious beliefs is initially problematic for the predictive processing account. But further development of the predictive processing framework suggests that we can satisfactorily explain this insensitivity. In particular, predictive processing can explain when high level beliefs influence representation of particular objects.

Finally, in the Chapter on *Persons* (Chapter 7), the sensory and dynamical aspects of person identification were discussed. These aspects of social cognition have only recently come into focus in cognitive science, and in this chapter I break new ground by discussing them in relation to misidentification syndrome. Accommodating these social cognition aspects in predictive processing will lead to more integrative account that can explain when a subject relies more on high level beliefs and the identification becomes more perceptual.

In every chapter, the hierarchical structure equipped with precision expectation played a pivotal and actually constitutive role—the higher level model "creates" contents at lower levels, where relevant sensory stimulation were not available at the time of experience. The hierarchical structure enables the brain to overcome lack of sufficient information, and in this sense, we live in a world partially created by our own imagination.

One common theme throughout the thesis concerned the *perceptuality* of each phenomenon. Using a definition for perception in which all models that are used to explain sensory stimulation are perceptual, I explained the five Ps occur in perception at least in some cases. A key strength of the predictive processing account is demonstration of the need of higher-level models that have belief-like characteristics even in daily perception. The need for higher-level models in predictive processing strongly suggests that higher-level models are a proper part of our perception.

To end my thesis—there are two kinds of "dance" that have played indispensable roles throughout the thesis. One dance is between high-level models and low-level models: high-level models exert top-down effects on low-level models. The other dance is between philosophical analysis and empirical science. All the chapters for the five Ps begin with philosophical analysis of the subject matter, which then sets the scene for, and is facilitated by, the predictive processing framework. In turn, science can benefit from philosophical analysis. I formulated some empirical predictions in each chapter. These empirical issues should be tested in future research. Now that one dance is over, new dance, led by empirical side, should start in the meantime.

Bibliography

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. *The Journal of neuroscience, 16*(23), 7678-7687.

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*(3), 396.

Association, A. P. (2013). *DSM 5*: American Psychiatric Association.

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex, 80*, 125-140.

Bauer, R. M. (1986). The cognitive psychophysiology of prosopagnosia *Aspects of face processing* (pp. 253-267): Springer.

Bayne, T. (2009). Perception and the reach of phenomenal content. *Philosophical Quarterly, 59*(236), 385-404.

Bayne, T. (2010). *The unity of consciousness*: Oxford University Press.

Bayne, T., & Spener, M. (2010). Introspective humility. *Philosophical Issues, 20*(1), 1.

Block, N. (1995). How many concepts of consciousness? *Behavioral and brain sciences, 18*(02), 272-287.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and brain sciences, 30*(5-6), 481-499.

Bonnet, C. (1760). *Essai analytique sur les facultés de l'âme, par Charles Bonnet.* . Copenhague: les frères C. et A. Philibert.

Boucenna, S., Cohen, D., Meltzoff, A. N., Gaussier, P., & Chetouani, M. (2016). Robots learn to recognize individuals from imitative encounters with people and avatars. *Scientific reports, 6*.

Briggs, F., Mangun, G. R., & Usrey, W. M. (2013). Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature, 499*(7459), 476-480. doi:10.1038/nature12276

http://www.nature.com/nature/journal/v499/n7459/abs/nature12276.html - supplementary-information

Brown, R. (2014). Consciousness doesn't overflow cognition. *Frontiers in Psychology, 5*(1399). doi:10.3389/fpsyg.2014.01399

Bullot, N. J. (2014). Explaining person identification: an inquiry into the tracking of human agents. *Topics in cognitive science, 6*(4), 567-584.

Burge, T. (1991). Vision and intentional content. In Lepore, E and van Gulick, R (Ed.), *John Searle and his Critics*: Wiley.

Campbell, J. (2009). Consciousness and reference. *Oxford Handbook of Philosophy of Mind*, 648-662.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition, 80*(1), 179-213.

Carruthers, P. (2003). *Phenomenal consciousness: A naturalistic theory*: Cambridge University Press.

Carter, G. G., & Wilkinson, G. S. (2013). *Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment.* Paper presented at the Proc. R. Soc. B.

Chadha, M. (2015). Time-series of ephemeral impressions: the Abhidharma-Buddhist view of conscious experience. *Phenomenology and the Cognitive Sciences, 14*(3), 543-560. doi:10.1007/s11097-014-9354-2

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*: Oxford University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *the Journal of Philosophy, 78*(2), 67-90.

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*: MIT press.

Clark, A. (2012). Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind*, fzs106.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences, 36*(03), 181-204.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7-19.

Cohen, J. (2002). The grand grand illusion illusion. *Journal of Consciousness Studies, 9*(5-6), 141-157.

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological science*.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in cognitive sciences, 16*(8), 411-417.

Corlett, P., Frith, C. D., & Fletcher, P. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology, 206*(4), 515-530.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adapted mind*, 163-228.

Dainton, B. (2006). *Stream of consciousness: Unity and continuity in conscious experience*: Taylor & Francis.

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association, 60*(3), 441-458.

Davies, M. (1992). *Perceptual content and local supervenience.* Paper presented at the Proceedings of the Aristotelian Society.

Dennett, D. (1991). *Consciousness Explained*: Penguin.

Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *The British Journal of Psychiatry, 157*(2), 239-248.

Evans, G. (1982). *The varieties of reference*: Clarendon Press Oxford.

Fazelpour, S., & Thompson, E. (2015). The Kantian brain: brain dynamics from a neurophenomenological perspective. *Current opinion in neurobiology, 31*, 223-229.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4*, 215.

Fish, W. (2010). *Philosophy of perception: A contemporary introduction*: Routledge.

Frege, G. (1948). Sense and reference. *The philosophical review*, 209-230.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. (2011). Embodied inference: Or I think therefore I am, if I am what I think. *The implications of embodiment (Cognition and Communication)*, 89-125.

Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology, 83*(2), 248-252.

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3*, 151.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Front. Psychol, 3*(130), 10.3389.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience, 3*(2), 191-197.

Geldard, F. A., & Sherrick, C. E. (1972). The cutaneous" rabbit": A perceptual illusion. *Science, 178*(4057), 178-179.

Gerrans, P. (2015a). All the Self We Need. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Gerrans, P. (2015b). Passivity Experience in Schizophrenia *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness* (pp. 325).

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron, 54*(5), 677-696.

Grush, R. (2007a). A plug for generic phenomenology. *Behavioral and brain sciences, 30*(5-6), 504-505.

Grush, R. (2007b). Time and experience.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences, 4*(6), 223-233.

Hayden, G. F. (1999). Alliteration in medicine: a puzzling profusion of p's. *BMJ, 319*(7225), 1605-1608.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural brain research, 32*(3), 203-218.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience, 3*(1), 80-84.

Hohwy, J. (2012). Attention and Conscious Perception in the Hypothesis Testing Brain. *Frontiers in Psychology, 3*(96). doi:10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The predictive mind*: Oxford University Press.

Hohwy, J. (2014). Elusive phenomenology, counterfactual awareness, and presence without mastery. *Cognitive neuroscience, 5*(2), 127-128.

Hohwy, J. (2015). Prediction Error Minimization, Mental and Developmental Disorder, and Statistical Theories of Consciousness. In R. Gennaro (Ed.), *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness* (pp. 293-324). Cambridge, Mass.: MIT Press.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687-701.

Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time (1893–1917)*. Netherlands: Springer.

Ichikawa, J. (2009). Dreaming and imagination. *Mind and Language, 24*(1), 103-121.

Jackendoff, R. (1987). *Consciousness and the computational mind*: The MIT Press.

James, W. (1890/2013). *The principles of psychology*: Read Books Ltd.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics, 14*(2), 201-211.

Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology, 24*(2), 175-219.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. *Scientific explanation, 13*, 410-505.

Kouider, S., De Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in cognitive sciences, 14*(7), 301-307.

Langdon, R., Connaughton, E., & Coltheart, M. (2014). The Fregoli delusion: a disorder of person identification and tracking. *Topics in cognitive science, 6*(4), 615-631.

Lau, H. (2007). A higher order Bayesian decision theory of consciousness. *Progress in brain research, 168*, 35-48.

Lau, H. (2008). Are we studying consciousness yet. *Frontiers of Consciousness: Chichele Lectures, 2008*, 245.

Lau, H., & Brown, R. (forthcoming). The Emperor's New Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations. In A. Pautz & D. Stoljar (Eds.), *Themes from Block*: MIT.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences, 15*(8), 365-373.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuroscience, 8*, 302.

Lee, G. (2014). Temporal experience and the temporal structure of experience. *Philosophers' Imprint, 14*(3).

Lerario, A., Ciammola, A., Poletti, B., Girotti, F., & Silani, V. (2013). Charles Bonnet syndrome: two case reports and review of the literature. *Journal of neurology, 260*(4), 1180-1186.

Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry?

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science, 245*(4919), 761-763.

Lycan, W. G. (1996). *Consciousness and Experience* (Vol. 72): Mit Press.

Mack, A., & Rock, I. (1998). *Inattentional blindness* (Vol. 33): MIT press Cambridge, MA.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*.

Martin, M. G. F. (2008). Commentary on Action in Perception. *Philosophy and Phenomenological Research, 76*(3), 674-681.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience, 5*. doi:10.3389/fnhum.2011.00039

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

McGinn, C. (1982). *The character of mind*: Oxford University Press Oxford.

Millikan, R. G. (1989). Biosemantics. *the Journal of Philosophy, 86*(6), 281-297.

Mole, C. (2011). *Attention is Cognitive Unison: An Essay in Philosophical Psychology*: Oxford University Press Usa.

Montague, M. (2011). The phenomenology of particularity. *Cognitive Phenomenology*, 121.

Nanay, B. (2010). Perception and imagination: Amodal perception as mental imagery. *Philosophical Studies, 150*(2), 239 - 254.

Newen, A. (2015). Understanding Others: The Person Model Theory. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group.

Noë, A. (2002). Is the visual world a grand illusion? *Journal of Consciousness Studies, 9*(5-6), 1-12.

Noë, A. (2004). *Action in perception*: MIT press.

Noë, A. (2006). Experience without the head. In T. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 411--433): Oxford University Press.

Noë, A., & O'Regan, J. K. (2000). Perception, attention, and the grand illusion. *Psyche, 6*(15), 6-15.

Oliva, A. (2005). Gist of the scene. *Neurobiology of attention, 696*(64), 251-258.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences, 16*(10), 504-510.

Prinz, J. (2012). *The Conscious Brain*: Oup Usa.

Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*(1), 127-158.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature neuroscience, 14*(12), 1513-1515.

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*: William Morrow New York.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87.

Rensink, R. A. (2000). The Dynamic Representation of Scenes. *Visual Cognition, 7*(1-3), 17-42. doi:10.1080/135062800394667

Rivolta, D., Palermo, R., Schmalzl, L., & Coltheart, M. (2012). Covert face recognition in congenital prosopagnosia: A group study. *Cortex, 48*(3), 344-352. doi:http://dx.doi.org/10.1016/j.cortex.2011.01.005

Rolls, E. T. (2004). A higher order syntactic thought (HOST) theory of consciousness. *ADVANCES IN CONSCIOUSNESS RESEARCH, 56*, 137-172.

Rosenblum, L. D., Smith, N. M., Nichols, S. M., Hale, S., & Lee, J. (2006). Hearing a face: Cross-modal speaker matching using isolated visible speech. *Perception & psychophysics, 68*(1), 84-93. doi:10.3758/bf03193658

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49*(May), 329-359.

Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. J. Flanagan, & G. Guzeldere (Eds.), *The Nature of Consciousness*: Mit Press.

Rosenthal, D. M. (2002). Explaining Consciousness. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (Vol. 46, pp. 109-131): Oxford University Press.

Rosenthal, D. M. (2005). *Consciousness and Mind*: Oxford University Press Uk.

Rosenthal, D. M. (2007). Phenomenological overflow and cognitive access. *Behavioral and brain sciences, 30*(5-6), 522-523.

Russell, B. (1921/2005). *Analysis of mind*: Routledge.

Schellenberg, S. (2010). The particularity and phenomenology of perceptual experience. *Philosophical Studies, 149*(1), 19-48.

Schlicht, T., & Newen, A. (2015). Kant and Cognitive Science Revisited. *Logical Analysis & History of Philosophy/Philosophiegeschichte und Logische Analyse, 18*.

Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review, 117*(2), 245-273.

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive neuroscience, 5*(2), 97-118.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences, 94*(7), 3408-3413.

Sherman, M. T., Seth, A. K., & Kanai, R. (2016). Predictions Shape Confidence in Right Inferior Frontal Gyrus. *The Journal of neuroscience, 36*(40), 10323-10336. doi:10.1523/jneurosci.1092-16.2016

Shimojo, S., & Nakayama, K. (1990). Amodal representation of occluded surfaces: role of invisible stimuli in apparent motion correspondence. *Perception, 19*(3), 285-299.

Siegel, S. (2012). Cognitive Penetrability and Perceptual Justification. *Nous, 46*(2), 201-222.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in cognitive sciences, 4*(4), 147-155.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649. doi:10.3758/BF03208840

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future.

Trends in cognitive sciences, 9(1), 16-20.

Soteriou, M. (2000). The particularity of visual perception. *European Journal of Philosophy, 8*(2), 173-189.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences, 23*(05), 701-717.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience, 8*, 666. doi:10.3389/fnhum.2014.00666

Strawson, P. F. (1974). Causation in Perception *Freedom and Resentment*: Methuen.

Teunisse, R. J., Zitman, F., Cruysberg, J., Hoefnagels, W., & Verbeek, A. (1996). Visual hallucinations in psychologically normal people: Charles Bonnet's syndrome. *The Lancet, 347*(9004), 794-797.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* (Vol. 282): Mit Press.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review, 121*(4), 649.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330-349.

Watanabe, T. (1995). Orientation and color processing for partially occluded objects. *Vision Research, 35*(5), 647-655.

Wittgenstein, L., Anscombe, G., & von Wright, G. (1967). Zettel Edited by GEM Anscombe and GH Von Wright. Translated by GEM Anscombe.

Woodward, J. (forthcoming). Scientific Explanation. https://plato.stanford.edu/archives/win2016/entries/consciousness-temporal/

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in cognitive sciences, 13*(4), 167-174.

Yovel, G., & O'Toole, A. J. (2016). Recognizing People in Motion. *Trends in cognitive sciences, 20*(5), 383-395.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in cognitive sciences, 10*(7), 301-308.

Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness). *Brain, 114*(2), 811-824.

Zimmermann, M. (1989). The nervous system in the context of information theory

Human physiology (pp. 166-173): Springer.