Third-person and First-person Perspectives: An Analysis of Thought-experiments

Timme Romberg

July 22, 2013

First supervisor: Janneke van Lith-van Dis
Second supervisor: Menno Lievers
Third reviewer: John-Jules Meyer

Studentnumber: 0477907
Master program: Cognitive Artificial Intelligence
ECTS: 60

Abstract

The clash between science and ordinary life has caused numerous discussions in particular about the mind. Out of these discussions about the mind arises a conceptual confusion. This confusion is caused by making no differentiation between conceptions belonging to the scientific domain, and ordinary day-to-day life. Such confusion is somewhat odd, considering that the clash is between the two distinct conceptions of the world where those specific conceptions belong to. These two conceptual domains have themselves been conceptualized in different forms. I will discuss three of them, namely: Nagels'objective point-of-view and subjective point-of-view, Sellars' scientific image and manifest image, and Dilthey's methodological dichotomy between Verstehen and Erklären. These dichotomies form the basis of my own description between the third-person perspective and the first-person perspective. With this framework I will be able to specify this conceptual confusion, show how such conceptual confusion arises, and finally analyze two thought-experiments (Mary's room and the Chinese room) which are illustrations of concrete examples of this conceptual confusion. By doing so I hope to provide conceptual clarity with respect to the (philosophical) discussion about the mind.

Contents

Preface

This paper is about conceptual confusion. I started thinking about conceptual confusion when I encountered very bright people that seemed to be unable to understand what I meant saying when I talked for instance about qualia or phenomenology. I tried explaining the difference between red as something physical and red as a qualitative characteristic of human experience. However, I was unable to communicate properly that there was such a difference (what I deem to be two distinct conceptual domains). It became clear to me that some *only* talk about red as a physical manifestation (e.g. the complex relationship between the wavelength of light and the visual information processing as it takes place in the brain of such a wavelength) and nothing else. Reflecting on my own ideas I realized that I usually talked about red as something phenomenological. I concluded that we were not talking about the same things, which made communication at best difficult and at worst meaningless.

My personal motivation for writing this paper is to have a framework within which such problematic communication can be placed and resolved. I also expect other people to have run in to the same problem; this paper should also help them. Of course efficient and clear communication does not stand on its own; it serves a purpose. Science as a social enterprise requires (proper) communication; if communication breaks down due to one not comprehending what the other has said this would have an effect on scientific progress. My intention is thus not to show that a purely physical or a purely phenomenological point-of-view is correct (or better) but that we should become more aware of the fact that people take different viewpoints (and by doing so prevent conceptual confusion).

Introduction

From this motivation I have formed my thesis: there are broadly speaking two distinct perspectives on the world and in particular the mind; the discourse surrounding the mind is influenced by people taking these perspectives. These perspectives lead to distinct discourse by virtue of consisting of distinct conceptual domains. Conceptual confusion arises when a concept belonging to one such domain is used as if it belongs to another. If such misplacement arises then this in turn has the effect of determining another distinct and distorted discourse. This confusion (and its effect) may not be noticed and when this happens the discourse remains distorted. Such unnoticed distortion lends itself to even more confusion (as not noticing such particular misplacement invites even further misplacement).

My goal will be to make this thesis plausible by describing these perspectives (chapter 1) and their relationship (chapter 2) and then, by having provided a framework in which to talk about these perspectives, show that these perspectives take a crucial role in the discussion about the mind and (more importantly) that not being properly aware of the distinction and relationship between these two perspectives creates conceptual confusion, which in turn determines the (content of the) discussion. This analysis will be done by taking a closer look at two philosophical thought experiments (chapter 3) using the third-person/first-person framework that will be described in chapter 1 and chapter 2 .

The construction of this framework will be done by investigating literature that describes dichotomies similar to the one that I have in mind. These dichotomies are: Nagel's subjective point-of-view and objective point-of-view (1.1), Sellars' manifest image and scientific image (1.2) and Dilthey's Verstehen and Erklären. Nagel's dichotomy consist of a particular point-of-view which is centered around a self (i.e. the subjective point-of-view) and a point-of-view which is no point-of-view in particular (i.e. the objective point-of-view). Sellars' dichotomy on the other hand consists of a conceptual image of the world which revolves around the concept that man has himself as a person (i.e. the manifest image) and one where he is no longer such a person due to the introduction of theoretical postulations (i.e. the scientific image). I will also discuss Dilthey's methodological distinction between a method to understand the world of general objects (i.e. Erklären) and the world of unique subjects (i.e. Verstehen).

All three of these dichotomies revolve around a similar conflict; namely, how one should see the work of science in relation to our ordinary day-to-day conception of the world. This problem is well known in several more specific forms. For example, how can free will exists in a natural deterministic world as described

by the natural sciences? What is the place of morality, art and the good life in view of a description of the world that has no inherent meaning or value. What is the relationship between body and mind? This thesis will primarily revolve around problems of that last type, problems that involve the mind. To be even more specifc I am concerned with the mind as viewed from the field of artificial intelligence.

A.I. as a scientific research project is an attempt to understand the workings of the mind by seeing the mind as some type of information processing unit, for instance a computer. Such a position leads to several philosophical problems. What is consciousness and if the mind is a computer then how would a computer be conscious? What does it mean to think, and how would a computer be capable of thought?

These problems arise because it is not intuitively clear how the mind is a computer. As a result there is a tension between the experience of the mind (and its workings) and the conception of the mind as a computer. This tension is of particular importance in the field of A.I. It is the topic of numerous discussion, there is a long standing debate whether or not we can we have an AI that fully functions as our mind. Consensus has not been reached but there seems to be a majority in favor of the idea that the mind is a computer (or similar information processing unit), that the mind can be understood as a computer and that one can have an A.I. that is fully functional (i.e. has a mind like ours). Nevertheless there is a recurring discussion about the relationship between the mind (as we know it) and information processing units. Within the framework that will be constructed in chapter 1 the former will classified as first-person conception and the later as a third-person conception. Discussions about A.I. are centered around the clash of these conceptions in particular and more generally their corresponding perspectives.

The relationship between first-person and third-person perspective will be discussed in chapter 2. There I will describe the role of models and metaphors in science and will show that they serve as a bridge between the two perspectives. I will do so by discussing the work done by Bailer-Jones, Brown and Lakoff and Johnson.

Support for A.I. is often found on the side that emphasizes the viability of a third-person perspective. Skeptics often emphasizes the first-perspective. These discussion often take place around a thought-experiment. In chapter 3 I wil discuss two of these thought-experiment, namely Searle's Chinese room and Jackson's Mary room. Both are prominent thought-experiments within the field of A.I..

I will analyze these two thought-experiments with the constructed first-person/-third-person framework and will conclude that there is indeed a conceptual confusion between the two perspectives.

1 First-person perspectives and third-person perspectives

My thesis relies on the idea that there are two different perspectives on the mind - a first-person perspective and a third-person perspective. This is in part an empirical claim; people take these perspectives in their real lives; it is not just an academic distinction. But because this is a philosophical paper I can not provide hard empirical evidence. Therefore evidence in favor of this dichotomy can only come in a different form. I am hoping that there is at least some shared personal experience with the anecdotes I have presented so far, and thus some shared intuition. The following discussion about different dichotomies which are similar to my own should not only serve as the starting point of making the third-person/first-person dichotomy explicit but should also strengthen this intuition further. This by itself is not enough; intuitions can be deceiving. In section 3 evidence will be presented in the form of an analysis of two thought-experiments. With the use of the framework that will be constructed in this chapter I will show what goes wrong when these thought-experiments are discussed. —It should then become clear that the third-person/first-person dichotomy is not only useful but also an accurate depiction (in a broad sense) of the different perspectives on the mind.

The dichotomy that I want to construct is broad. Its parts are generalizations of more specific token perspectives. There is no such thing as *the* first-person perspective. Both perspectives are distinct categorical types of perspectives. These token perspectives share a structure but not their specific content. Each of these token perspectives may themselves consist of further subtypes. I will use the terms (the) first-person perspective and (the) third-person perspective to denote what is true for all the specific tokens of perspectives that fall under one such type of perspective. My focus in this paper is on this broad distinction between the first and third-person perspective, but I will sometimes refer to a more specific token perspective. The same holds true for the other dichotomies that I will start discussing in this chapter; all are such generalizations.

These dichotomies are in order of appearance: Nagel's objectivity and subjectivity (1.1); Sellers' manifest image and scientific image (1.2); and Dilthey's Verstehen and Erklären (1.3). I will use these generalizations in order to construct my own dichotomy in section 1.4. These are three different dichotomies. I do not equate them and consider them to only be similar to one another to the extent that they have certain relevant commonalities. *They are in this sense part of the same cluster of dichotomies.* By looking at these dichotomies which

share the (broad) view that the world can be studied and understood in two distinctive ways, I will be able to construct a dichotomy of my own (1.4).

Some might consider these dichotomies wholly different, or at least too different to draw any parallels between them. I do not think this correct. This will be clarified later on in section 1.4. But I will give some preliminary indication why they are similar now. The most obvious similarity is that all three describe two different ways of looking at the world; all three dichotomies provide two different conceptions of this world. (This is not really true because the distinction between Verstehen and Erklären is methodological and does not *directly* describe two different conceptions of the world. But I do think it does so indirectly. This will be argued for in section 1.3.) They differ with respect to what exactly these conceptions are and how they are related. Nonetheless, one of those conceptions (the subjective point-of-view, the manifest image and the perspective that Erklären prescribes) within their own dichotomies always corresponds to a more everyday kind of view towards the world, in which a more personal and less abstract view is present. Conversely the other elements of these divisions move away from that everyday mode of thinking. Each of these dichotomies is about opposite *world-views*[1]. I will consider these three dichotomies as a part of unspecified *cluster* of dichotomies. They may not have been intended to be placed within such a cluster and their reasons for construction may have also been different. Nonetheless, for the purpose of my thesis they can be examined as belonging to this cluster. Their commonalities will be discussed in section 1.4 along with their relevant differences. I will first discuss each dichotomous world-view separately.

1.1 Objectivity and subjectivity

In this section I will discuss Nagel's dichotomy between the subjective point-of-view and the objective point-of-view. My aim in this section will be to give an overview on Nagel's thoughts about this dichotomy. I will do so by giving a rough sketch of what its parts are (1.1.1) and then clarify these further by discussing how they are related to one another both in terms of origin and opposition (1.1.3 and 1.1.4).

[1]Or partial world-views if one considers a union or a 'double vision' of both to be a world-view

1.1.1 What are the subjective and the objective points of view?

The objective point-of-view is a view of the world not from any *place* in particular but from nowhere in particular [27, p. 208]. This results in a perspective that is not bound by bias or anything considered to be personal. Such an objective point-of-view is thus an *abstraction* away from anything personal, what Nagel calls "[A] transcendence of the self" [27, p. 209]. This transcendence of the self should not be confused with taking someone else's particular perspective. Transcendence of the self as an abstraction means transcending above your self *and any self*. The other (non-abstracting) form of transcendence will be (briefly) discussed when I integrate Dilthey's Verstehen and Erklären (1.4.2).

Objectivity in this sense should not be confused with the objective truth. Nagel: "Objectivity is a method of understanding. It is beliefs and attitudes that are objective in the primary sense" [28, p. 4]. The objective point-of-view can thus be described as having an attitude towards the world with certain goals and beliefs. These beliefs consist of what these goals are and how to accomplish them. One of these goals would be to understand the world (which is roughly the same as obtaining objective knowledge). How does one do this? "[W]e step back from our initial view of [the world] and form a new conception[.]" [28, p. 4]. Such an attitude is in part an ontological *position* from which one looks at the world. Such an ontology entails using a certain methodology and specific corresponding epistemological content.

That old conception comes from a perspective that is composed of all those things personal. This subjective point-of-view is the personal and human perspective we take when living our ordinary lives. It is the conception of a being (which is the subject itself) and the world in terms that are *unique and dependent on its own particular point-of-view* [42, p. 2]. It is an egocentric perspective - the world as viewed from each one's own particular point-of-view. In that sense, the subjective point-of-view is context dependent; it always depends on *where* an individual finds himself in the world[2].

These two points-of-view stand apart conceptually but are still related. The objective point-of-view is a detachment from the subjective point-of-view. This detachment takes the form of an abstraction. This kind of abstraction leads to a non-personal (and thus less-contextual) body of knowledge (i.e. possession of the objective truth). It is due to this detachment from the subjective point-of-

[2]I should add that Nagel considers there to be a gradual scale between objectivity and subjectivity; the above stated definition of subjectivity as egocentric should then be placed on the far subjective side of this scale. However, for practical reasons Nagel idealizes this scale in to a non-gradual dichotomy. I will continue to discuss Nagel's dichotomy as he has; which is as if the idealization is a truthful non-idealized representation.

view that the objective point-of-view forms a conception of the world and the self in terms that are not unique or dependent on any particular perspective [42, p. 2].

Such an ontological view of the world should correspond to a descriptive body knowledge which fits with this ontological point-of-view. This is an ideal and may never be reached. Nonetheless, both knowledge and epistemic content are shaped by this ideal which is inherent to the objective point-of-view.

1.1.2 What is the correct point-of-view?

Both perspectives provide a picture of the world which is either incomplete or faulty *as seen from the other point-of-view*. I prefer examining the dichotomous world-view cluster from *within* each dichotomous part because it ensures a position of neutrality. Throughout this paper I will stay neutral with regards to the superiority of either the first-person perspective and third-person perspective and the dichotomies I use to construct them. I will do so by maintaining that both perspectives have a distinct conception of what the world is, what kind of knowledge one can have of it and how to acquire it; I will refrain from dealing with truth values. To be more precise: I will not be concerned with truth values as *independent* of a perspective *even if there is such a thing*.

My position of neutrality is in part equivalent to Nagel when he says that "The correct course is not to assign victory to either standpoint." [28, p. 4]. But Nagel does assign *defeat* to the objective perspective. For instance; ".. not all reality is better understood the more objectively it is viewed."[28, p. 6]. And: "[T]here are things about the world and life and ourselves that can not be adequately understood from a maximally objective standpoint."[28, p. 7]. Nagel has a preconceived notion of how things are. The idea of 'reality' or life and self as they truly are as independent of either of these two perspectives (in Nagels case objective and subjective, in mine: first-person/third-person) means choosing sides. I do not assign victory *nor defeat* to either perspective. In order to make the previous judgment that Nagel makes with regard to reality one has to know what reality is. If I would say what reality is then I would have choose sides which means not being neutral[3].

Let me elaborate on this point. Nagel with respect to the origin of the objective point-of-view: ".. only the *supposition* that we and our appearances are part of a larger reality makes it reasonable to seek understanding by stepping back from appearances[.] [emphasis mine]" [28, p. 4]. This supposition is part of

[3]This is not exactly true; I will make claims about what these perspectives are and how they are related.

the basis of the ontological structure of reality *within the objective perspective*. Only by looking from outside the objective point-of-view would one consider this to be a *supposition*. Within it it is not just a supposition, from within it is considered to be true. Furthermore, 'true' in this case refers the objective truth (embedded within the objective perspective). To be clear, when I describe certain conceptions as being distorted or faulty I do not intend to say that they are faulty/distorted but rather that they are seen as such from the other perspective.

My view on what such a world-view dichotomy (like Nagel's) entails (concerning the elements and relationships within it) also differs. In the case of Nagel this is because of the distinct nature of our goals. My goals is a practical goal to uncover conceptual confusion whereas Nagel's goal is more ambitious. Nagel is primarily concerned with "[h]ow to combine the perspective of a particular person inside the world with an objective view of that same world." [28, p. 3]. I am primarily concerned with the confusion that arises out of the attempts to resolve such a problem (the problem of trying to bring together conceptually distinct perspectives while not realizing the nature of this conceptual distinction).

1.1.3 What is the origin of the objective point-of-view?

During the course of our human existence we find distortions within our subjective view of the world [27, p. 197]. We realize that not everything is what it appears to be. The subjective point-of-view is thus considered to be distorted and requires compensation/correction. Such correction requires the use of appropriate methodological tools. The uneasiness with our subjective perspective motivates us to find such tools to correct these distortions. When found, we use them to remodel this misrepresentation of the world.

Finding such tools is problematic. Reaching an objective conception of the world requires tools that are part of our human cognition[4]. This cognition was considered to be the cause of distortion. Therefore, such tools as a part of our cognition may be distorted as well. As a result, one could always question the objective nature of the corrective equipment that it has found within itself [27, p. 208]. However, one thing is known; that it is the subjective point-of-view that is distorted. Therefore whatever tools are the least subjective would be the most appropriate to be used to make an objective judgment [27, p. 208].

We then search for tools that are the least biased and the most impartial, which is that part of ourselves that depends least on the contingencies of the self. Such

[4]Such an objective conception is an epistemological conception of the world; the ontological conception was already present

contingencies would include logical biases, perceptual illusions/limitations, personal tastes/feelings, territorial drift but also certain intuitions, commonsense ideas about the world, faulty correlations, etcetera. As a result of this demand we acquire tools that are the most accessible to others. It is in this sense that there is a process of *abstraction*, with respect to methodology, when thinking about the objective's relationship to the subjective (i.e. a depersonalization of the subjective point-of-view by removing the personal contingencies). Such methodological abstraction can thus be characterized as a continuous search for increasingly more abstract tools (away from the self) to use for critique of the self and its own subjective point-of-view [27, p. 208-209]. I agree with Nagel on this and the idea that objectivity requires this form of abstraction will be one of the main ingredients for the construction of my own dichotomy (1.4).

This process of methodological abstraction differs from but runs parallel to the process of ontological abstraction. The methodological abstraction is a shift from reliance on immediate experience to the use of logic, mathematics and other tools that do not rely on the contingencies of the self. But one can only view the world with the use of such objective if they are applicable to it. For example, in order to describe the world in mathematical terms the world has to be seen as adhering to mathematical principles which may not be how the world appears through a subjective point-of-view. As a result the ontology of the world has to shift as well.

This can be clarified with Nagel's idea of how an objective *physical* conception of the world is formed [28, p. 14]. Initially what is physical is what appears to us through sensory perception. We first realize that our perception is caused by the effects of some external world on our (physical) body. The next realization is that those external activities also effect other external physical things. We then come to the conclusion that there is an external physical world which has to be understood as such. We have to conceive of a world that is independent of us seeing it *and how we see it*. This goes against the egocentric nature of the subjective perspective and in this sense we can speak of an *ontological distortion* (as seen from the objective point-of-view).

This is the reason why the subjective perspective is often considered to be erroneous. It is considered to be erroneous *from that objective point-of-view* because a subjective perspective is directly dependent on our (direct) perception of the world. The process of perception is initially seen as standing apart from the world that is observed (as Nagel points out); it is not a part of the external world; it is not caused by it.

It seems to me that if perception is caused without appearing to be so then there

might also be other (possibly more specific) mechanisms that we are not aware of. This leads to the conclusion that the subjective point-of-view is not only false but also incomplete[5][6]. Later on when discussing Sellars something similar will be noted namely in the form of the postulation of theoretical inperceptibles (1.2).

1.1.4 Why is there still a subjective point-of-view?

Once there is an objective point-of-view we have gained a point-of-view next to the old subjective point-of-view. This new perspective should have replaced this subjective point-of-view according to the objective point-of-view. However, the use of objective tools has only created a new conception of the world *without* replacing the old one. This runs contrary to the objective's goal of *correcting* the old distorted point-of-view. In actuality, the need for correction has caused two separate points of view. The ideal to have *one* full and complete non-distorted point-of-view of the world has not disappeared; it has (only) not been fulfilled.

The subjective perspective has remained, according to Nagel, because some of our subjective experiences cannot be accounted for by an objective point-of-view [27, p. 210]. These aspects of the subjective can not be clearly judged as real or not. For instance, with these tools at our disposal we may come to the following general objective description of mental states[7]. Mental states are causal relations between stimuli and behavior; they function as causal intermediaries between the stimuli input and behavioral output [2]. This view (which is a functionalist view) on mental states provides an impersonal and comprehensible account of mental states. Such a conception of mental states is appropriate from an objective point-of-view but is not from a subjective point-of-view. This account of mental states is detached from one's own personal subjective experiences of those mental states; it is the result of correcting those mental experiences to the point where we can consider them to be objectively real (in opposition to our subjective notion of mental states).

To put it briefly, the methodology as prescribed by the ideals of the objective point-of-view, to view the world from an impersonal point-of-view with tools that make it possible to do so, has resulted in different kinds of abstraction (i.e. different token point-of-views within the objective point-of-view). It has

[5] And therefore also false in its claim that it is complete.

[6] I realize that this 'reason' for the supposed erroneous nature of the subjective point-of-view is incomplete. But if I, Nagel or someone else would have given such a complete reason then this paper would not exist.

[7] I am calling these descriptions 'objective' because they are descriptions that have been made by taking an objective point-of-view. I do not mean to say that they are 'objective' in the sense that they exist, even if they do.

also led to an abstraction moving away from not just a methodology derived from personal experience and an ontology consisting of experience but also the experience itself; subjectively observed properties are no longer considered to be real. That does not mean that those experiences have disappeared, they still seem just as real *in their subjective form* as before [27, p. 201]. The less one relies on one's own subjective experiences the more one approaches an objective account of reality, yet the less this objective account corresponds to the way that things appear to oneself [27, p. 209].

1.2 The manifest image and the scientific image

Nagel's distinction between the subjective and the objective is similar to a dichotomy made by Sellars. This dichotomy draws a distinction between what Sellars calls the manifest image and the scientific image. These images are also two distinct perspectives on reality [35, p. 5]. I will discuss this distinction in the following sections. Its similarity with Nagel's dichotomy will be elaborated on in section 1.4.

Sellars deliberately uses the term 'image' to describe the two elements in his dichotomy. By calling them images he avoids saying which one (or both) is true (or false). Like Nagel Sellars does not *intend* to judge either image as true or false (even though Nagel does). It is for this reason that Sellars talks about 'images': "[B]y calling them images I do not mean to deny to either or both of them the status of 'reality'." [35, p. 5]. An image as a metaphor for a conception also eludes to that image being 'taken' from a place which is distinct from its counter-part. The image is thus the result of taking a certain *point-of-view* (cf. Nagel). There is thus a manifest point-of-view and scientific point-of-view corresponding to a manifest image and a scientific image. Those large scale conceptions of reality/the world also consist of smaller more specific conceptions which I will call manifest or scientific phenomena. For example, we can speak of the manifest mind as a specific conception formed from a manifest point-of-view and as part of the manifest image. This relationship will be elaborated on in section 1.2.2 but I will first discuss what the scientific image and the manifest image are.

1.2.1 What are the manifest image and the scientific image?

The 'manifest image' is described as that which is formed by investigating observable phenomena and finding correlations between them *without any theoret-*

ical postulations of unobservable entities/mechanisms/processes etcetera[8]. The manifest view on the world can thus be characterized as consisting of those phenomena that one is (directly) aware of and the correlations found between those phenomena. "[T]he manifest image is the image of [...] ordinary moral, social, and interpersonal life." [3, p. 301]. Naturally those correlations are not fixed and vary across time and space, this means that several manifest images have existed. Within each such image the person as a whole is the object of analysis [3, p. 301]. That is not to say that a manifest image is unscientific; it could (and does) use the scientific method[9].

The manifest image is what an individual finds himself with when becoming aware of himself and of the world around him [35, p. 6]. People have always found themselves to be in the world and so have always had an image similar to our current manifest image because we have always been aware of the world in a similar way[10]. Nonetheless, the manifest image of today is not the original image. The original image is the way people viewed the world around them the first time people were aware of the world they found themselves in.

The manifest image is a (purely) conceptual image (and so is the scientific image). This is an important difference with Nagel; his subjective point-of-view does not merely result in the *conception* of free will (as in the manifest image) but also includes the 'raw' experience of free will from which this conception has been made and also where it refers to. The manifest *conception* can not include pre-conceptual raw 'experience'. This does not mean that it is not supported by it. The manifest image and the world as experienced do share a close connection which I will argue for later on.

The original image is also a purely conceptual image which means that it has not been formed by conceptual thinking. Its inception is spontaneous and not a gradual conceptual development. In this original image all things had motives, beliefs and intentionality [35, p. 10]. Clearly this is not our manifest image of the world any longer; plants for instance are not (or rarely) seen as having a personality, they are no longer considered to be persons. This shows that the manifest image is not static. The manifest image is a refinement (or adjustment) of the original image and could still undergo further refinement. Our current

[8]Some (e.g. Dennett) argue that beliefs/motives/intentionality are theoretical postulations themselves.

[9]Whatever that may be.

[10]This is an empirical assumption but it is embedded within the manifest image. If the manifest image sees ancient man as not experiencing himself as a person then he would not be anything like present day man which contradicts the manifest conception of ancient man as a person. For example, characters in ancient Greek literature are thought of as persons because they show intentions, feeling, conscious awareness of the world, etcetera, just like we do therefore they must have been persons and must have experienced the world like we do.

manifest image is distilled to the point where all but persons themselves have been depersonalized [35, p. 10]; it is this current manifest image that is the topic of this section and that will later be analogous to the first-person perspective (1.4).

As I have mentioned before, the manifest image does not postulate any unobservable theoretical entities, processes or mechanisms; the scientific image does. The scientific image seeks to explain the correlations that have been found in the manifest image. The scientific image, by seeking such explanations, allows in its need to explain the features of the manifest image, that certain unobservable entities/processes are postulated [35, p. 10]. Such postulations have, in the course of the scientific image's development, increasingly little to do with persons and the intuitions about them. This idea, that the scientific image requires unobservable postulations to account for the correlations made between observable phenomena, will be an important part of the construction of the third-person/first-person dichotomy (1.4). Additionally one can already see some similarities between one of the other important ingredients that I have discussed in the section about Nagel's objective and subjective points of view (1.1). There I mentioned abstraction (as in the use of methodological tools that are the least dependent upon self) as playing an important role; such abstraction leads to a point-of-view that seemed alien to one's own experience of the world. Postulation of theoretical unobservable entities has an analogous effect - they are not a part of our manifest understanding of the world (1.4).

1.2.2 Subjective experience as part of the manifest image

I would now like to further clarify the relationship between Nagel's subjective point-of-view and Sellars' manifest image. I have previously described Nagel's subjective point-of-view as a place from which one starts to conceptualize the world and that the manifest image is the resulting conceptualization of the world. We could then say that a point-of-view is the ontological starting point of investigation and the image is the resulting body of knowledge. In this section I will show how the manifest image as a body of knowledge *itself* is related to the subjective point-of-view as the ontological starting point of investigation. In order to do so I will start by giving a more in depth description of the relationship between original image and the current manifest image.

The original (manifest) image is "[T]he framework in terms of which man came to be aware of himself in the world." [35, p. 6]. That is to say, the image in which man *conceptualized* himself as man and thus also became the man that he is (because that conceptualization is a part of being man). The (current)

manifest image is a refinement of that original image. To be more specific the current manifest image stems from a number of adjustments based upon previous adjustments going back to the original image.

Nevertheless, the current manifest image is an adjustment of the initial stage of conceptual thinking (i.e. the original image). The original image is man's first conceptualization the world; Sellars asserts that within this conceptualization man did not only conceptualize himself as a person but also everything around him; everything was *categorized* as a person. By perceiving the world in such a way man was able to make correlations between these persons. For example, man saw how the wind (as a person) interacts with the leaves and branches of a tree (as a person). From such a personal point-of-view the following scenario could then occur: man made the *empirical* conclusion that the wind *wanted* to push the tree away and that the tree takes a defensive stance by leaning back. By having an ontology of the world in which everything is a person man can only possess knowledge about the interactions of elements in the world in terms of personal interactions.

What does it mean to be a person? With person Sellars means "a way of being a person"[35, p. 10], but not necessarily a person like you and me are persons. In the original image we would be human-persons (what we still consider to be correct) and trees would be tree-persons (what we now consider to be false); they are not the same kind of persons as we are but were still some form of person. This is longer the case.

Nevertheless, the original image and the manifest image do not only share the same ontological structure in which persons can exist, they also share an ontology of entities/things that could be categorized as a person. In the original image everything was a person but that leaves open the question what one considers to be a thing. It seems to me that man's specific pre-conceptual interaction with the world is what lead to its specific original image of the world (with specific entities being persons). For example, man interacted with trees as whole entities. Man's behavior and perception is evolutionarily adapted in such a way to allow a certain type of (pre-conceptual) interaction. This interaction with trees is what made trees as a whole to be conceptualized as people.

The entities that were considered (or the entities that could have been conceptualized) to be persons remain as secondary (non-person) objects precisely because they were entities which could be interacted with in a primal behavioral, experiential and pre-conceptual way before being conceptualized as persons. Their conceptualization as entities does not rely on them being persons (or being conceptualized as such). Which has as a result that when they were

no longer considered to be persons they remained part of that image as non-person entities because nothing changed pre-conceptually (i.e. raw experience). A certain (subjective) point-of-view must thus have remained in place and is prior to conceptualization.

To make this more clear it is best to discuss Sellar's distinction between habitual and intentional behavior. Habits are behaviors that are repeated without being aware of them; they are nevertheless part of being a person. Non-person entities do not have habits; a tree (as a non-person) does not drop its leaves every autumn because of habit; it is merely *caused* to do so. This is contrasted with the acts of people: "most of the things people do are not things they are caused to do"[35, p. 13].

Why am I discussing this? Because it shows that acting with intention (and not by mere cause) is an essential part of being a person. Such non-causality of the acts of a person does not entail free will because the experience of thoughts that come from nowhere does not mean that they are of a free kind[11]. Nevertheless, such non-causal and thus intentional acts and thoughts are one of the links between the manifest image and the subjective point-of-view; intentionality is subjectivly experienced (before being conceptualized) and the manifest image conception of a person is based on that subjective aspect of being a person.

One could still ask whether man *has* intentionality because of the experience or the conception of intentionality and thus whether the conception of intentionality is prior to the experience of intentionality or vice versa. The manifest perspective is the proper perspective to take in order to answer this question, because only within the manifest perspective is there such intentionality.

The manifest image does not see itself as merely a conception of the world; both the image as a whole and its parts are truthful. It is for this reason that within the manifest image there is intentionality prior to its conception. In other words the manifest image relies on subjective point-of-view[12].

I have pointed out twice how immediate experiences are essential to, but not a part of, the manifest image. 1) The continued existence of depersonalized entities as the same (yet differently categorized) entities and 2) the deliberate acts of people requiring (non-conceptualized) experience prior to the conceptualization of intentionality. I have done so in order to show that the manifest image is in part a conceptualization of subjective experiences. It are those experiences (as well as subjective conceptions) that Nagel also discussed.

[11] Rather it points the opposite; even phenomenologically do thoughts seem to occur without any control.

[12] From the outside this may well be a mere postulation.

1.2.3 The clash of images

The endorsment of the the manifest as real (in one form or another) is called perrenial philosophy. The scientific image has also been endorsed as real. I would now like to discuss what happens when these two images come in to contact with each other. This will serve as a prelude of the clash between third-person and fist-person perspectives. In this section I will in particular focus on the *self-containment* of these images.

The endorsement of perennial philosophy of the manifest image is embedded within the manifest image itself. The conceptual elements and their relationships within the manifest image are seen as providing a true representation of the world. Without this 'self-affirmation' the manifest image would be inconsistent; it would not be able to make any correlations between (what it has to consider to be) real entities. Because the manifest image is the image that one finds oneself with during ordinary life it is also implicitly endorsed when one is living this ordinary life.

Such endorsement of the manifest image as real leads to conflict with the scientific image: "[The scientific image] purports to be complete image"[35, p . 20] as well. And thus: "[f]rom its point-of-view the manifest image on which it rests is an 'inadequate' but pragmatically useful likeness of reality." [35, p . 20]. Man is man by virtue of having a (manifest) conception of himself. And thus, in order to preserve himself man has to preserve the (manifest) image of himself. This identification is threatened by the scientific image in which man is *no longer categorized as a person.* Even more, there is no person-category within this scientific image.

Nonetheless: "The scientific image is supported by the manifest world."[35, p. 20] and tries to provide further understanding of manifest phenomena (both introspective and extrospective). The difference between the scientific image and the manifest image is "between that conception which limits itself to what correlation techniques can tell us about perceptible and introspective events and that which postulates imperceptible objects and events for the purpose of explaining correlations among that which is directly observed."[35, p. 20]. The scientific image is methodologically dependent on the manifest image. Furthermore, from a manifest perspective such a foundational connection also means that the scientific image is substantively dependent on the manifest image, which means that "the scientific image cannot replace the manifest without rejecting its own foundation."[35, p. 21]. The scientific image itself has to dismiss this specific position because it *also* claims to be the the complete and true image of the world.

What this means is that each image contains within it some representation of the other image as well as itself and a relationship between its own image and the other image. And so a critique or analysis of these image can come from within its own and from the other image. In other words, there is a critique on the image which is the image of that image. This may be a self-reflective image or an external image. For instance, from the scientific perspective the manifest image is considered useful and similar to reality but still inadequate, or: "[The perennial tradition's] attempt to understand the achievements of theoretical science in terms of this framework, subordinating the categories of theoretical science to its categories"[35, p. 19]. *Such positions are analogous to the core of the conceptual confusion that I would like to discuss.*

As we will see this type of critique is often misplaced; such critique does not just come from an other perspective is also assumes that its own ontology and its corresponding epistemology are correct and can be used to formulate *successful* critique (directed at the other image).

This is problematic for two reasons: 1) this critique is directed at the image of the other image. 2) success is determined by the other image adapting (or adopting) a more suitable position which means dropping its own claim to reality. Not only does the other image not feel compelled to change, it can not change.

Thus, although the criticism is intended to be directed at the other image it is actually *self-contained*. This kind of (usually unnoticed) self-containment is one of the core parts of the confusion between third-person and first-person domains which I will discuss later on.

I would now give some examples of such self-contained criticism. Sellars gives us three solutions that are available when confronted with the clash of images. The relationship between imperceptible theoretical particles/objects is either one of 1) identity 2) abstraction or 3) illusion [35, p. 26]. Position 1) is reductionism where manifest objects are composed of smaller theoretical entities. Within position 2) one can find fictionalism where theoretical entities are seen as mere fiction serving as at best symbolic or abstract representations of manifest objects. On the most radical side of position 3) one can find eliminative materialism where the manifest world is seen as (completely) hallucinatory.

In order for reductionism to be accepted by the manifest image it would have to accept the ontology of the scientific image. For example, one could argue that it is possible to reduce phenomenological colour to physical light waves but there are no theoretical light waves *within the manifest image* to be reduced

to[13]. The same applies to persons; even if we rephrase the scientific ontology as not stating that there are no persons then there would still be theoretical entities which can not exist as a part of the manifest image.

Another way out is by stating there is no such thing as a manifest color; that its qualities are mere appearances [35, p. 27]. This leads us to position 3) and is then countered with the knowledge argument: manifest objects exist because we know they exist (here Sellars refers to Moore). In other words, it goes against common sense to deny the existence of, in our case, phenomenological color. However, this is simply restating part of the ontology of the manifest image, that perceptual objects have perceptual qualities. Therefore the knowledge argument "operates *within* the framework of the manifest image and cannot *support* it."[35, p. 28].

The only options that then remain are stating that theoretical light waves do not really exist (and are only symbolic abstract representations), which is position 2; light waves *exist* as something separate from color as a sense-field. In the last case this means that "if the human body is a system of particles, the body cannot be the subject of feeling and thinking" [35, p. 29]. This body is known as a physical body because it is conceptualized as such. Simultaneously, "we know what thinking is without conceiving of it as a complex neurophysical process, therefore, it can not *be* a neurophysical process" [35, p. 30], we know that thinking is such and such because it is conceptualized as such. Bringing these two points together means recognizing that it is not due to their *true nature* that, for instance, identifying neurophysical states with (manifest) mental states is problematic, but rather that problems occur due to their *conceptual nature*.

However, dualists like Descartes have concluded from this that "[n]either sensation or feeling or conceptual thinking could in this sense be construed as complex interactions of physical particles, or man as complex physical system" [35, p. 29]. This means more than the more obvious statement that man does not directly observe himself to be composed of theoretical entities; it also means that man does not identify himself with the observable body. *Thus my initial point that the mind/body problem is a conceptual problem caused by comparing the manifest mind with a scientific body becomes mute. One is still left with the manifest mind/manifest body problem. Which brings up the question whether the manifest image is dualistic in its very essence*[14].

By presenting dualism like this I am reformulating the way dualists have ap-

[13]This argument mirrors Sellars argument, his argument is about certain manifest properties that do not exist within the scientific image.

[14]This is not Descartes dualism but it is the right reformulation considering that such dualism is based on a conceptual confusion.

proached the mind/body problem. Let me exemplify. According to Sellars dualists are prepared to say that "a *chair* is really a system of imperceptible particles" and that man *is* not. Such a claim about reality points out that they do differentiate between manifest man and scientific man; they are *only* talking about manifest man. But they are also talking about the scientific body; this is the conceptual confusion that I want to discuss. Nevertheless as I discussed earlier, this does not say anything about the mind/body problem as a problem *within* the manifest image. Seen as such, dualism is simply one of the many perennial philosophies which endorse the manifest image as real but is in conflict with other perennial philosophies about its specifics.

1.2.4 The scientific image and the philosophy of mind

In order to understand the correlations in the manifest image, new entities and/or processes have been postulated which are deemed appropriate for doing so; such postulations created the scientific image. Our attempt to understand the correlation of the brain (and the rest of the body) and the (manifest) phenomenon mind would be an example of such theoretical postulation. This (manifest) phenomenon mind is the mind that we find ourselves to be equipped with - it is the mind we are aware of. These correlations are observed correlations between the observed mind and observed body. A (manifest) point-of-view only presumes body/mind dualism to be dualistic to the extent that they are experienced and observed as distinct entities, but since such experience is preconceptual it does not entail some form of dualism. Whatever the case, studying these correlations may then involve the postulations of unobservable theoretical entities/processes. Including such postulation means that the mind is no longer conceptualized as it was in the manifest image.

The study of human behavior (as it is correlated to stimuli) does not require such postulations. We find that certain behavior is correlated with certain stimuli and we can build a theory around these findings without postulating any hidden mechanisms. The study of the correlation between stimuli and behavior remains a manifest form of science, as long as both behavior and stimuli are only observed and thus does not involve any theoretical postulation.

Does this mean that behaviorism is a manifest form of science? There are several forms of behaviorism [17]. First, those that do not deny the existence of the mental but only refuse explanation in mental terms are still part of the manifest image. This type of behaviorism does not *use* any mental events, observed or unobserved. Only when a behaviorist claims that the mental events/states that we are aware of do not exist would his form of behaviorism not be part of the

manifest image. However, it does not become part of the scientific image either. If we follow Sellars' requirement that theoretical postulations are required to be part of the scientific image, then the denial of mental states itself does not make such a position a part of the scientific image. The scientific image is an expansion of the manifest image; it can not change the manifest image itself. A type of behaviorism that changes the manifest image is neither part of the scientific image nor of our current manifest image. It is as a (possible) future refinement of our current manifest image. The reason being that it is those states that are considered to be a part of our being as a subject; they are part of what it means to be a person. The removal of such states means a further depersonalization[15], a depersonalization of man himself[16].

Such a non-depersonalized view on the mind sees the mind as non-decomposable. The non-decomposed mind might not explain the manifest correlation found between stimuli and behavior. Further explanation requires decomposing this manifest mind. Several options then become available; we could study the brain directly (which we have found to be correlated to the manifest mind) or use some kind of a priori theorizing about what the mind would consist of. This shift from behaviorism to cognitivism was made possible by dispelling behaviorism's paradigmatic taboo of not looking inside the black box (that which makes it possible for input to become its corresponding output). The removal of this restriction leads to the acceptance of postulating internal states/processes/representations composing the black box [13, p. 497]. This shift should be understood as an ontological shift and will also be discussed in the section about Verstehen and Erklären (1.3).

Such a decomposed view of the mind does not correspond to the world we find ourselves to be in, our manifest image. From this point-of-view the person and its features are non-decomposable, they are basic [42, p. 3]. Scientific progress (specifically within the scientific image) has led to increasingly more depersonalization, which in turn induces more tension between the personal view (i.e. manifest image) and impersonal view (i.e. scientific image) [42, p. 7].

1.3 Erklären and Verstehen

The previously discussed manifest image and subjective point-of-view are not necessarily unscientific. In actuality one can categorize distinct branches of

[15] With 'depersonalization' I am referring to a process analogous to the removal of trees from the category of person.

[16] I am not sure what a manifest image would be where people are no longer seen as persons, because in our current manifest image people are all that are left to be persons.

science as adhering to either scientific image or the manifest image (or an objective or subjective perspective). Similar categorizations have been made since the beginning of the scientific revolution (e.g. what kind of science this or that science is) [23]. Such categorizations can result in philosophers making certain methodological prescriptions.

One of those philosophers was Dilthey. His methodological prescription is best expressed by his statement that "nature we explain, man we understand" [11][17]. Dilthey describes the human world as mentalistic in which the acts of people are not caused by blind (deterministic) causal laws but by human motivations. The causal laws on the other hand do the Erklären of the natural world, which is a world of objects. Because the world, for the natural sciences, is a world of objects it *should* then also be understood differently from the world of subjects, the human world. The social sciences are the ones that are concerned with this world of subjects. Such an ontology the world (as consisting of subjects) requires its own methodology in order to gain any knowledge about such a conception of the world. It is a methodology that does not concern itself with acquiring knowledge about the world conceptualized as a world of objects. This methodology revolves around the concept of Verstehen.

Verstehen should be understood as grasping human beings by focusing on *specific* subjects in a *specific* context. It considers each subject as a distinct unique subject. One is then always left with a part of such a unique subject that can not be generalized. Erklären does not consider the world to consist of unique entities and all entities can thus be generalized. It is in this sense that Verstehen is more focused on context.

Grasping the world of objects involves creating a distance from the human and personal world. This form of distancing would be the wrong approach when trying to grasp the world of subjects. Erklären , the methodology of the natural sciences, is not the proper way to grasp this human world. Instead, Dilthey would argue in favor of staying inside this human world. This entails actively taking human experience as a subjective phenomenon and to use it to understand subjects [12, p. 146]. By using the subjective/human point-of-view we would be understanding the human world by being a part of it. This would mean taking someones particular and specific subjective point-of-view. Such an approach requires imagining what it would be like to be a particular person in a particular context. By doing so we recreate the mental states of such a specific individual in our own mind. The mental states recreated consist of beliefs and intentions. Furthermore the actions of subjects should be understood as

[17] Although a more literal/accurate translation would be: "Nature we explain, the life of the soul we understand."

being chosen without restrictions as contrasted with the deterministic and law-abiding world of objects. In short, the aim of human sciences is to understand the world of subjects; this is done by understanding the subjects "*immediate,* lived experiences [emphasis mine]" [23, p. 25].

Although the distinction between Erklären and Verstehen is methodological, and is thus prescriptive about how different branches of science should grasp the world, it also makes the ontological distinction between a world of objects and a world of subjects. Such ontological assumptions may not necessarily imply that there actually is a world of objects and/or a world of subjects (although it could be argued for) but it does imply that we, *in practice*, can look at the world in these two different ways. In other words, the distinction between Verstehen and Erklären also states that there could be two distinct epistemologies about the world. A methodology only makes practical sense if we are actually capable of using it to produce knowledge; if a methodology does not lead to an epistemology it would be useless.

The methodological distinction between Verstehen and Erklären thus entails two distinct epidemiologies, each based on distinct ontological assumptions. The perceived completeness and accuracy of these epistemologies gives rise to an increased assurance of the ontological assumptions underlying these epidemiologies. The knowledge/understanding compromising these two distinct perspectives gives further justification for the ontological assumptions made. That does not mean that we should reduce talk about ontology to epistemology but just that in practice epistemology is involved with maintaining and restructuring its underlying ontology. I am not equating epistemology with ontology in one way or another. However, looking down on an epistemology, one should see that ontology and epistemology are interwoven; knowing an ontology produces a certain epistemology which in turn justifies having and using such an ontology.

1.4 The first-person perspective and the third-person perspective

So far I have discussed three separate but related dichotomies. These dichotomies are not the same but do share certain commonalities. I will now use these dichotomies and their commonalities to create a dichotomy of my own, between the third-person perspective and the first-person perspective. I will do so by drawing parallels and observing similarities between the dichotomies that have been described. These associations will serve as the basis for the distinction between the third and the first-person perspectives.

Before starting to describe the dichotomy that I have in mind I should add that a somewhat more subtle depiction of the opposition that is found in science (and philosophy) is possible. However, my aim is not to give a perfect depiction of this opposition. Each dichotomy discussed shows in its own way that such an opposition is indeed there and tries to give a formulation of it. My aim is to make clear the implications of such a dichotomy and the consequences of not being aware of it (3). This can best be done by going straight to the core of this opposition without specifying all the subtleties that may be involved; the topic is already complex enough and does not necessarily require any more clarification of it in order to serve the stated purpose.

I will reuse both the concepts and the terminology that I have discussed in the previous three sections. Such reuse of terminology should not be confused with the exact meaning of the original terminology - this new dichotomy is built from the previously described dichotomies and is as a whole separate but still derived from them and thus analogous to them.

1.4.1 What is a perspective?

Broadly speaking one could give two definitions of a perspective: 1) A way of looking at something 2) The place from which one looks at something.

These two definitions are interrelated and I will use both descriptions to describe what I mean by a perspective. Definition 2) corresponds to the idea of a point-of-view in the sense that it is a point from which one views something. In the cases of Nagel's point-of-views these places to view the world. In that sense they were ontological point-of-views. For the purpose of this paper I will not restrict a perspective to only being point-of-view. The perspectives that I would like discuss also include an epistemological image (c.f. Sellars) and a methodology (c.f. Dilthey).

The reason why I can include a method as part of a perspective is because the place from where one starts to look at something (e.g. point-of-view) determines how one looks at something (e.g. method). The reason why I include epistemological images is because the place from which one observers determines what one is going to see (which is also dependent on method). All three components of a perspective are thus interrelated.

A point-of-view in its most concrete form is a physical place in both space and time that an individual takes in order to observer something (e.g. a tree). The perspectives that I am discussing are not such spatial perspectives, and neither are their points-of-view. This is true even though the first-person perspective

that I will discuss in the next section is an abstract generalization of such particular spatial points-of-view. I will call the first-person perspective and the third-person perspective 'world-perspectives'.

I will not discuss the the validity of these perspectives. I will discuss these perspectives as being self-reinforced and self-contained. With self-reinforced I mean the validity that a point-of-view, image and method derive by their interrelatedness. For example, A point-of-view produces an image which is deemed as correct but this would also mean that the point-of-view is correct.

Another more specific example, the token third-person perspective of chemistry (which is itself a part of the token third-person perspective of the physical sciences) has as its basic unit the atom. The ontological point-of-view of chemistry contains the knowledge that the world consists of atoms. If the world consists of atoms and the goal is to bring into focus this world of atoms then a methodology is also required. This methodology in turn provides a concrete image of the world in which two oxygen atoms bond with a carbon atom to form carbon dioxide. Such knowledge of carbon dioxide does not merely presuppose atoms it also confirms their existence because knowledge of carbon dioxide entails knowledge of (non-specified) atoms. Or in other words, the ontological conception of the world as atomic both determines and is found within the specific conception of carbon dioxide.

Another feature of these world-perspectives is that they are also *self-contained*. Meaning that the perspective as a whole stands on its own and supports itself. This is related to the self-reinforced aspects of a perspective. A perspective derives validity due to the interrelatedness of its components and can therefore support itself. This has as a consequence that any critique on that ontological point-of-view will be evaluated by that same ontological point-of-view. In short, each perspective stands on its own and will evaluate any critique on it from within that perspective. This important because critique from outside such a perspective is often given as if it is not. This is one way in which conceptual confusion can manifest itself. In chapter 3 this will be explored further.

1.4.2 Abstraction and postulation

By now I have only made some short remarks about the specifics of the world-perspectives that I would like to discuss. I will start by elaborating on the third and first-person perspective by first making some clarifications with respect to abstraction and spatial points of view. I am adopting Nagel's notion of the objective point-of-view as a view from nowhere. This means something

more then saying that the third-person perspective is non-spatial; it would not differentiate with the first-person perspective if it was not. The third-person point-of-view is not just non-spatial in the sense that it is an abstract description of commonalities of several token points-of-view, it is also non-spatial in the sense that those token points-of-view are also non-spatial. The third-person perspective is an abstract description of several views-from-nowhere. This differs from the first-person perspective which, although is itself non-spatial, is not an abstraction of other non-spatial views.

For example, the first-person point-of-view describes the (ontological) commonalities shared by the token-perspective of some person looking at a tree close by and another person looking at a tree from far away. These are both spatial points of view while *the* first-person perspective is not. The third-person perspective is non-spatial either but so are its token points-of-view.

The ontology of the first-persons perspective (which is its point-of-view) is a world that consists of persons. When discussing Sellars I pointed out that to say that some entity is a person is equivalent to assigning it certain person-like features, for instance intentionality and free will. Furthermore, the possibility to assign such features is dependent on the immediate experience of these features. It are these immediate experiences that were the basis of the *formulation* of Nagels subjective point-of-view. When discussing Nagel it was made clear that only from a particular subjective point-of-view one could have immediate experience. For someone to be aware of their own experiences is to be that particular someone. Such features are thus fully dependent on having a particular subjective point-of-view. Having such a point-of-view means first (introspectively) observing oneself as having person-like features and *then* being able to extend them towards others. The ontology of the first-person perspective as a world consisting of persons is thus dependent on the observation of immediate (introspective) experience (which is itself another experience) and must therefore also acknowledge the existence of these experiences. Sellars' manifest image as an image revolving around persons is thus strongly related to Nagel's notion of the subjective point-of-view revolving around one's own personal experiences.

Nagels objective point-of-view *aims* at being a perspective that is (fully) independent of any personal perspective (1.1). It considers the subjective point-of-view to be flawed. It believes that the world can not be understood from a subjective point-of-view because it is inherently flawed. The world does not consist of those things that become apparent through immediate experience. This is similar to the scientific image attitude towards the manifest image. The scientific image does not accept that the manifest image is correct. Within the manifest image man sees himself as a distinct entity contrasted with all other

entities. By postulating theoretical entities/processes (with respect to man) one could then say that how he acts and how he thinks is similar to all those others entities. He would be just another entity with no special (person-) features.

The way in which the objective point-of-view aims to have an epistemology that does not depend on immediate experience (1.3) is by abstracting away from it. Such abstraction is done by providing generalizations of the world (and the mind). This component of the third/first-person dichotomy called *abstraction* is taken from Nagel (1.1). The third-person perspective is an abstraction away from the methodology used in the first-person perspective. This first-person methodology consists of using our naturally equipped perceptual equipment (this includes introspection) and is partially described by Dilthey (1.3).

In effect, by relying, on its own personal non-abstracted perceptual equipment the first-person perspective does not use any theoretical postulations of unobservable entities/processes/mechanisms to understand the world; it is based only on those things/phenomena that one can be aware of (1.2). The second component of my dichotomy (next to abstraction) is taken from Sellars and is called *postulation* (1.2). Postulations are used to explain the correlations found between manifest phenomena; these postulations are not part of the first-person perspective. Such postulations intrude on what is considered to be real from the first-person perspective, they are nonexistent within the subjective experience of the world.

Postulation and abstraction are not the same. They do however have some important commonalities. Abstraction makes use of tools that are not dependent on the contingencies of the self. Postulation is such a tool; postulation inserts things that we are not aware of, that are not dependent on us being aware of them, and so they too are abstract and impersonal. By postulating such abstract entities we make first-person phenomena decomposable. Whereas prior to postulation an entity only consisted of certain subpart which were also first-person phenomena. By using theoretical postulation one could then decompose these subparts even further by stating that they consist of smaller theoretical parts. I will later explain how such abstract entities are still linked to a first-person conception of the world (2) (which is responsible for the conceptual confusion).

One might argue that manifest phenomena are decomposable; for instance, a tree can be decomposed in root, trunk, branches and leaves. It is true that manifest phenomena can be decomposed into smaller *manifest* phenomena and still remain part of the manifest image. The key point is that this kind of decomposition does not use any theoretical postulation of unobservables. The manifest image consists of manifest phenomena that can be decomposed into

smaller part as long as they are *also* a part of the manifest image. These parts are also manifest phenomena and thus not unobservable theoretical postulations. Henceforth I will use 'non-decomposable' as shorthand for 'non-decomposable using unobservable theoretical postulations' and 'decomposable' as shorthand for 'decomposable using unobservable theoretical postulations'.

The objective point-of-view, as Nagel describes it, sees itself as a correction of the distorted subjective point-of-view (1.1). The scientific image is the body of knowledge which was the result of explaining the correlations made between manifest phenomena (1.2). These are complementary goals. While searching for explanations of manifest correlations one could discover that such correlations are actually incorrect. For instance, when wanting to explain a certain manifest entity one might find that it did not exist in the first place. Similarly, in order to correct a first-person perspective, the use of decomposition, postulation and explanation provides extra support in doing so. I will thus see the goal of the third-person perspective as the need to explain (using abstract postulations) *and* correct (from its own perspective) what is found in the first-person perspective.

1.4.3 Subjects and objects

The third-person perspective, as described so far, does (methodologically) two things differently with respect to the first-person perspective: abstraction and theoretical postulation of unobservables. This methodology follows from the aim of the third-person perspective to provide a non-distorted and more complete epistemology. However, to be able to use this methodology and provide a certain epistemology a distinct ontology is also needed. In this section I will use Dilthey's distinction between Verstehen and Erklären (1.3) to elaborate on this distinction.

Verstehen is a prescriptive methodology. It says that the world should be understood as a certain kind of world. This world is a world in which people are persons. This world of persons is the manifest world. It can only be understood by not abstracting away from the personal point-of-view in which humans ordinarily find themselves - such understanding does not allow postulation or abstraction. Human beings should be understood as particular individuals within a particular context, it requires imagining being someone else *as he finds himself in the world*. Such an imagining can only consist of things that are directly accessible to our individual being (i.e. manifest phenomena). This is the form of transcendence that Nagel said was opposite to objective transcendence, to remove oneself from oneself *and* any self, which does require abstraction.

We will see that imagining being some other entity contributes to the confusion between first-person and third-person perspective (3). There can not be any imagining of what it is like to be a person within the third-person perspective. Such imagining requires persons to be manifest persons. Such persons can not be decomposed or abstract entities; such entities do not exist within a third-perspective. As a result one can not use the methodology of Verstehen from a third-person point-of-view. Verstehen requires persons to be manifest persons (i.e. non-decomposed and non-abstracted persons); it requires human beings to be subjects (and not objects).

Verstehen requires an ontology in which the world is comprised of subjects, integrating this in to third-person/first-person framework: the first-person perspective rests on an ontology in which the world consists of subjects. However not everything in a first-person perspective is a subject; we are not talking about Sellars' original image where everything was subjective (1.2). The first-person perspective consists of subjects to the extent that we consider certain manifest phenomena (specifically people with manifest minds) to be subjects; such a view does need an ontology of a world of subjects and (manifest) objects. A third-person perspective on the other hand needs an ontology of a world of (decomposable) generalized objects. In turn it provides a general, lawful and depersonalized account of the world.

1.4.4 Two conceptions called 'mind'

So far I have described two perspectives on the world. This also means that there are two different perspectives on the mind. These two perspectives each have a distinct conception of the mind (and its features). One is the mind as we are aware of it, a non-decomposable mind. The other is the mind as a decomposable object. The subjective conception of the mind I will call the manifest mind and its objective counterpart the objective mind. This does not mean that these conceptions are not related; the third-person perspective on mind is structurally related to the first-person perspective of mind by use of metaphor, which will be discussed later on(2).

The first-person image of the mind is formed by our immediate experiences. It is a commonsense concept. This conception of the mind and its features is basic and non-decomposable. It can not be decomposed and is not composed of any unseen theoretical entities/processes; there is nothing hidden outside our awareness of the mind that is considered to also be a part of the (manifest) mind. This brings up the issue of the subconscious mind. The above description of the first-person image does not allow for existence of the subconscious (within

that first-person image). If all conceptions of the world are based on immediate experience and can not consist of any postulation of imperceptibles then the subconscious as something that can not be immediate experienced and thus requires postulations does indeed not exist within the first-person conception of the world. Whatever the subconscious is it is not a part of the first-person conception of the mind.

What I call 'the objective mind' is a generalization of several (third-person) theories about what the mind is. These theories have in common the ideal of giving an objective description of the mind, a decomposable and abstracted description of the mind. We know that a mind exists as a phenomenon by virtue of our manifest conception of the mind. The objective mind is the result of trying to explain what the manifest mind is; it is a consequence of trying to figure out how it works and how it is related to other phenomena (for instance our actions (i.e. outputs)). Doing so requires *not* viewing the manifest mind as non-decomposable. Such a conception of the mind (as decomposable) allows us to break apart the mind with the use of theoretical postulations. The objective mind is thus a modification of the manifest mind. Such a modification leads to a new separate conception of the mind. Although the initial intention may be to improve on the manifest mind, we end up with a new conception of mind, that does not correspond to our manifest mind. Within the first-person perspective this adaption is a distortion and thus not accepted as real. Within the third-person perspective this is indeed an improvement.

As long as one uses a manifest conception of the mind to create an objective conception, it would also be (at least partially) dependent upon the old conception of the mind. Furthermore it is by virtue of us having a manifest conception of the mind that we know that there is a mind; that is not to say that the phenomenon mind is dependent upon the conception of mind, just that we *know* that there is a mind by virtue of our manifest conception of the mind. We therefore need to modify our conception of the manifest mind in such a way that we end up with some objective conception of the mind.

It is important to avoid seeing this divide as implying an ontological or metaphysical divide; to say that there are (broadly speaking) two distinct perspectives on the mind both highlighting different aspects of the mind does not imply that there actually are different substances or properties at work, *despite the ontological assumption that there are*. This can best be elaborated on by using Northoff and Mosholt's analysis of Searle's defense against the claim that he is a property dualist [29][34].

Searle claims that mental states (as experienced) are irreducible to third-person

mental states. By doing so he positions himself against materialism (at least those materialists that are also reductionists) [29, p. 593]. However, this view does not have any ontological consequences. To say that the epistemic difference implies an ontological difference is called (by Northoff and Mosholt) the epistemic-ontological inference. Consciousness can be *causally* reducible to the brain but the first-person ontology can not be reduced to a third-person one [34, p. 58]. Searle claims it is the ontology that is irreducible while Northoff and Mosholt are saying that these views are epistemologically irreducible.

Within my framework the ontologies used are described as ontological assumptions that are needed to provide a certain knowledge-base. From this description it is more appropriate to view the epistemologies as irreducible or else one requires viewing the assumptions made as actually correct, thus leading to (property) dualism. From this view the distinction between the two perspectives only has contingent conceptual consequences.

Although we are able to experience our mental states (as first-person mental states), we do not have the ability to perceive the physical states responsible for it; we do not perceive our brain states in any way. This is called the autoepistemic limitation [29, p. 590]. The neglect of this limitation is what makes possible the epistemic-ontological inference. Physical brain states can only be seen from this third-person perspective and mental states can only be seen from a first-person perspective. In this sense the mind-body problem has an epistemic cause not an ontological one because there is no need to draw any ontological inferences just because we lack certain (innate/natural) perceptual tools for doing so. This means that a first-person perspective on the world creates an epistemology based on an ontology that comes with these perceptual tools; it does not mean that this ontology is correct.

1.5 The story so far; dichotomies about the mind

What I have discussed so far are three different dichotomies; I have used them to make clear my dichotomy between the first and third-person perspective (1.4). I have done so by using both the ideas behind these dichotomies and their terminology. The first person-perspective is the world we ordinarily find ourselves to be in. It consists of manifest phenomena and correlations between them; this is how the world is understood from a first-person perspective (1.2). In it, we consider human beings to be subjects, which are thus understood as such. Such an ontology about the world requires its own methodology in which we imagine specific subjects in a specific context (1.3). We use our own first-person understanding of ourselves in order to understand the world (the

world containing subjects). The third-person-perspective aims at explaining the manifest correlations made (1.2) and providing a view of the world that is not dependent on any subjective point-of-view (1.1). It does so by making use of tools that are the least dependent on the self, by removing all possible subjective elements. By abstracting away from the self it hopes to find an objective account of the world (1.1). This requires seeing human beings as objects and not as subjects (1.3). Manifest correlations are made between manifest phenomena. These phenomena are perceived from a subjective point-of-view and thus are (potentially) flawed (1.1). By abstracting away from this subjective reference point we allow ourselves to think of these phenomena as not being what they appear to be. These manifest phenomena can then be decomposed; they can consist of things that we are not aware of (1.2). This allows the postulation of theoretical unobservables. With these theoretical postulations we can give an objective account of the manifest phenomena and the correlations between them (1.2). As a consequence, a different methodology is used which allows a non-subjective dissection of the world, which results in a general description of the world (which is considered to be a world of objects) (1.3).

2 Scientific models as metaphors and its consequences

In section 1.4 I explained what the first-person and third-person perspectives are; one of the main conclusions was that both had a different conceptualization of the world (and the mind in particular (1.4.3)). In this chapter I will discuss how these perspectives (in particular their images) are conceptually related by the necessary use of metaphor. This will make clear why they are conceptually distinct and why confusion between the two perspectives can arise (2.3). I will do so by first discussing models as metaphors (2.1), then incorporate such a view of models/metaphors in the previously constructed third-person/first-person framework (2.2). Using this expanded framework I will be able to explain why conceptual confusion between first-person conceptions and third-person conceptions can arise (2.3). Cases of such conceptual confusion will be examined in chapter 3.

2.1 Scientific models and metaphors

In the following sections I will discuss what scientific models are (2.1.1), when they work (2.1.2) and discuss their crucial role in science (2.1.3). I will then describe what metaphors are (2.1.4) and finally show that models are metaphors (2.1.5).

2.1.1 What are models?

In order to describe what scientific models are I will use a characterization of models by Bailer-Jones. I will later describe how this characterization would fit in to the third-person/first-person framework (2.2.1). Bailer-Jones gives the following definition of a model: "A model is an interpretive description of a phenomenon that facilitates access to that phenomenon" [1, p. 108]. These interpretations may involve idealization or simplification of the target phenomenon. Models can also shrink or enlarge the objects and processes one is studying, by doing so aspects of phenomena that are *not directly observable* can be made explicit [1, p. 109]. I should add that such a shirking or enlarging happens on a descriptive level; I am not claiming that the actual ontological object is shrunk or enlarged. But I will later on argue that models can re-conceptualize first-person conceptions into third-person conceptions; in other words, provide an account of phenomena that fits with a particular ontological view of the world.

All the types of models discussed provide partial access. They are partial because they can only highlight parts of the phenomenon and therefore provide partial descriptions. But what does it mean to provide access? Bailer-Jones: "[P]roviding access means giving information and interpreting it, and expressing it efficiently to those who share in the specific intellectual pursuit." [1, p. 109]. Thus, providing access involves more than just providing information about a phenomenon; it also involves providing an interpretation of this information and being able to *efficiently* communicate this interpreted information [1, p. 109]. In this sense, models provide *common* access to a phenomenon. These two discussed features of models (providing a common access point and the ability to make unobservable entities/processes apparent) are of importance with respect to the third-person/first-person framework because these aspects of models make possible the re-conceptualization from first-person to third-person conceptions as will be discussed in section 2.2.1.

Models can take several different representational forms, which are not mutually exclusive [14]. There are the previously mentioned scale models (which are those models that shrink and enlarge objects) as well as idealized models (which simplify objects) but there are also analogue models (models that work by virtue of picking out similarities) and phenomenological models (models that do not postulate hidden mechanisms). None of these models effect the ontology of a phenomenon itself but they do provide a description which may or may not fit a particular ontology (for instance, a third-person point-of-view).

My primary focus will be on analogue models which are not phenomenological models. I will later on show that such models are tools that are used to re-conceptualize first-person conceptions in to third-person conceptions. For now I will only explain why I will not be concerned with phenomenological models.

Third-person conceptions differ with their related first-person conceptions by the inclusion of things that are not directly observed. Postulation of such things require models (this will become clear later on in section 2.2.1). These models can not be phenomenological models because such phenomenological models do not postulate hidden mechanisms. Phenomenological models can thus not function as a bridge between first-person conceptions and third-person conceptions[18]. Furthermore, my primary focus will be on analogue models because of the prevalence of the analogue model of the mind as a machine within the field of A.I. and philosophy of mind.

I will use Bailer-Jones' description of models as: 1) providing an interpretation, 2) providing common access, 3) providing partial access, 4) making explicit hid-

[18] At least with respect to theoretical postulation

den mechanisms. As I explained not all models have feature 4) but the models that I am discussing have this feature. Such a view of models is not compatible with the first-person point-of-view. These models are part of the methodology of the third-person perspective because they are used to re-conceptualize first-person conceptions in to third-person conceptions which are only deemed ontologically appropriate within that third-person perspective. For instance, the first-person perspective can not accept the description that models make explicit hidden mechanisms because there are no hidden mechanism within such a perspective.

In conclusion, models as tools are part of the third-person methodology and the description given of them is part of a third-person image (e.g. it is how they can be known within the third-person perspective). I will henceforth use 'model' to denote these models unless specified otherwise.

To say that models only provide partial access can be interpreted (in light of the third-person/first-person framework) in two different ways: 1) non-phenomenological models provide one part of complete access and phenomenological models the other part or 2) within the third-person perspective models can only provide partial access (e.g. each distinct non-phenomenological model provides a distinct (partial) third-person image). Both interpretations are problematic within the third-person perspective. 1) is problematic because it violates the third-person point-of-view of the world as independent of experience and being decomposable in to hidden mechanisms. 2) is problematic because it entails that the third-person image can never be complete and the third-person perspective this third-person perspective does sees itself as potentially complete. Interpretation 1) can not be correct because the models that I am talking about are part of the third-person methodology. Such models as used from a third-person point-of-view can thus not provide any first-person access. Interpretation 2) will be discussed in section 2.2.3.

2.1.2 When do models work?

Models work when they can provide the previously discussed access to a phenomenon, but what is required of a model to provide such access? To answer this question I will focus on what Watt calls the intuitive appeal of models. I will use his description of this intuitive appeal and combine it with the previously discussed characterization of models by Bailer-Jones.

I will first discuss Watt's view on the intuitive appeal of models. It should be noted that Watt is primarily talking about models (and metaphors) about

the mind. I will later show how his ideas can be generalized to include other phenomena.

According to Watt there are two kinds of connections between the mind and a model. 1) as a system of descriptions (as how Bailer-Jones described models) and 2) as having "intuitive psychological appeal"[40, p. 52]. Intuitive appeal is "a reflection of how well people can see the model as something that could be psychological" [40, p. 52]. What he means is that we are drawn to models of the mind when they seem like a proper description of the mind *as we know it*. Within the first-person/third-person framework we could describe this mind-as-we-know-it as a first-person conception of the mind. There is such intuitive appeal by virtue of having such a mind. Watt: "[Having a mind] colors the whole of cognitive science, acting as a continual pressure on the metaphors that we use." [40, p. 52]. In the context of the third-person/first-person framework this means that we are *familiar* with the mind (as a first-person conception) and look for models that reflect this familiar mind. Models about the mind that have intuitive appeal reflect in some way our first-person conception of the mind.

This description of intuitive appeal can be generalized to also include models of other phenomenon. We can reformulate intuitive appeal as the reflection of how well people see a model as something familiar; intuitive appeal can be expressed as the (intuitive) detection of (a high amount of) *similarity* between a model and something *familiar*.

The higher this similarity the lower the effort required to intuitively grasp a model because it is similar to something already known. What does it mean for a model to be intuitively grasped? Models where we know what they are about without making explicit everything that they imply are intuitively grasped.

Such grasping is needed because providing a full description of what a model implies requires too much time and effort for the scientist (in practical terms); there is too much information contained within a model for such information to all be explicated. Furthermore, such an explication may not even be possibly because a scientist may not have (nor ever have) such a clear (linguistic) explication of the model. Here I am (in part) reiterating Brown when he says that: "[T]he theorist cannot, and need not, make explicit all the implications of the model he is exploiting" [4, p. 82]. Luckily, we know what a model is about without needing to make everything about it explicit due to its similarity with something that is commonly familiar.

We learn from this two related things: 1) a third-person conception of a phenomenon can not deviate too much from its root first-person conception or else

it will not be intuitively grasped 2) the more familiar we are with this root first-person conception the less a model will deviate from this conception because the intuitive dissimilarity is then more easily spotted.

According to Morrison and Morgan "[m]odels are often used as instruments for exploring and experimenting on a theory that is already in place" [25, p. 19]. Access is not an end product of scientific work, the access provided by a model is used for exploratory reasons. Therefore, even if common access is provided when such access can not be readily intuitively understood exploration becomes difficult. In conclusion, good models are intuitively grasped. We can thus say that (good) models have intuitive appeal because they are intuitively grasped.

We can thus also say that models make certain things implicitly clear and that which they make implicitly clear can then be used to explicitly explore the target phenomenon. Such aspects of a good model are all prior to the accuracy it provides (or *can* provide by exploration). Thus the initial acceptance of a model does not depend on the accuracy of its description but upon its intuitive appeal which consists of intuitive grasping of a model. The accuracy of a description is something that is worked out; it is determined by exploring the consequences of a model as to how it relates to reality (i.e. gathered empirical data). Such consequences can not be determined unless we have at least some intuitive grasp of the model, which enables us to explore it. This intuitive grasping is then in part responsible for the accuracy of description. In other words, a good model works because it is understood prior to explication (i.e. it is intuitively grasped).

2.1.3 What role do models have in science?

I would now like to propose that models are a bridge between the first-person and third-person perspectives. I have already mentioned that the first-person and the third-person perspective are related in the sense that the third-person perspective is both an abstraction (it does not deal with particular perspectives or the related first-person experiences) and a further elaboration of the first-person perspective (first-person correlations are expanded upon with the use of theoretical postulation). I will now show how models (and later on I will equate these with metaphors) make this relationship possible. I will do so by first discussing Morrison and Morgan's depiction of models as mediating instruments and than integrating parts of this description in to the third-person/first-person framework in section 2.2.1.

According to Morrison and Morgan models can be seen as autonomous entities, meaning that they are independent of both theory and world [25, p. 8]. Yet

they include both; models consist of both theory and worldly data. By doing so they provide an *intermediary* between the two. Additionally, models consist of things that are not part of the original object of investigation; their construction involves both data, and world, and other elements. This observation by itself does not make models crucial for science. However, because of their independent construction they can also function independently; they provide an instrumental function that neither theory nor data can have [25, p. 18]. They are used to help build a theory, to explore the implications of a theory and can also limit the domain of abstract concepts [25, p. 20].

For the purpose of this paper I am primarily interested in the mediating roles of models between world and theory because of how this mediation is related to the bridge between first-person and third-person conceptions. If a third-person conception consist of theoretical postulations then these postulation have to be somehow linked to the world itself. This is what models can do with their intermediary function. This does not mean that all models function as a bridge between first-person and third-person conceptions. Models are however needed as a bridge between first-person perspective and third-person perspective because third-person conceptions consist of theoretical postulations that require an interpretation which is grounded in that which already understood.

The importance of discussing the role of models in science should in the context of this thesis be understood as the role of models in the third-person perspective. I do not equate the third-person perspective with all forms of science but do categorize all forms of science that deviate from the first-person perspective by means of abstraction and theoretical postulation as part of the third-person perspective. The role of models in science as Morrison and Morgan can thus be understood as the role of models with respects to the sciences within the third-person perspective as they mediate between *unobserved theoretical postulation* and world.

What Morrison and Morgan call the world is that what is empirically observed. They do not talk about first-person and third-person perspectives and therefore do not specify what they mean with world and therefore also not what is empirically observed. I am concerned with the confusion between first-person and third-person perspective and therefore interpret 'world' as first-person conception of the world (even though models do not only mediate between theory and first-person conception of the world). In the next two sections I describe models as types of metaphors and then this bridging function will become more explicit.

2.1.4 What are metaphors?

I will make use of Lakoff and Johnson's description of metaphors in order to establish what metaphors are. By using their work it will also be possible to provide argumentation that our view of the world is largely metaphorical.

Before continuing I would first like to clarify the difference between metaphorical concept, metaphorical expression and metaphor as a tool[19]. In the last sense 'metaphor' means the use of a certain psychological instrument to understand the world. I will call this metaphor-as-tool. The actual use of metaphor-as-tool produces a metaphorical concept. A metaphorical concept is the resulting conceptualization by means of metaphor-as-tool of some phenomenon. A metaphorical expression is an explicit linguistic expression of a metaphorical concept. When I talk about metaphors I will be talking about metaphorical concepts unless specified otherwise.

A metaphor, according to Jaynes, is "the use of a term for one thing to describe another because of some kind of similarity between them or between relations to other things" [20, p. 48]. This is not an accurate description of a metaphorical *concept* because terms are not concepts. It is however similar to a more accurate description of metaphors given by Lakoff and Johnson: "The essence of metaphor is understanding and experiencing one kind of thing in terms of another."[21, p. 5].

One thing that both Jaynes and, Lakoff and Johnson share is that they perceive metaphors as essential aspects of our understanding the world. What Lakoff and Johnson mean by that is that metaphors shape our conceptual system. I will agree with them on this point. Numerous examples in favor of this idea can be found in "metaphors we live by" by Lakoff and Johnson and the case-studies in chapter 3 can also be seen as support for this position.

The previous description of a metaphor can be construed as a specific kind of the basic structure of metaphors. Lakoff and Johnson describe this basic structure as X IS Y. X is that which is to be understood which they call the target, and Y is that which is used to understand which they call the source. I would like to add that this source, which is used for understanding, must itself be understood first. The source has to be familiar in order to *determine* similarity by some individual. It is for this reason that we can describe the use of metaphors as a methodological tool which uses previously established knowledge to construct something new (cf. models and familiarity).

[19] Both 'metaphorical concept' and 'metaphorical expression' are taken from Lakoff and Johnson.

An example of a metaphorical concept would be ARGUMENT IS WAR [21, p. 4]. Such a metaphorical concept consists of several different sub-conceptions. Both the main metaphorical concepts and its sub-concepts can be expressed in different ways. For instance, ARGUMENT IS WAR can be expressed as "argument is war". In a war one has strong or weak positions and thus the ARGUMENT IS WAR metaphor entails that one can also have weak or strong positions regarding an argument. Such a sub-conception can be expressed as: "Your position is weak".

I will use Lakoff and Johnsons' description of metaphors as understanding (and experiencing) things in terms of others things and having a structure X IS Y.

2.1.5 Are models metaphors?

My previous talk about metaphors may have already implicitly shown the relationship between models and metaphors. I will now make this relationship explicit. I will show that scientific models and metaphors have a similar function, work in similar ways and that they both require the same kind of intuitive appeal. This will later be useful in order to make explicit the role of models as bridge between first-person perspective and third-person perspective because we can link the role of models in science as mediating between theory and world (Morgan and Morrison) with Lakoff and Johnson's notion of metaphorical concepts as well as its structural implications.

Models have as a function the understanding of one thing in terms of another; this is equivalent to what metaphors do. Metaphors provide understanding of a target with the use of a familiar source. Metaphors thus provide additional understanding about the world by making use of what has already been understood. They do so by focusing on particular similarities between source and target. The same applies to models, a model provides understanding about a target phenomenon by drawing similarities. In both cases are the similarities drawn partial and in both cases we rely on a certain intuitive appeal. In conclusion, scientific models are metaphors. This does not mean that all metaphors are models. This is in accordance with Bailer-Jones' view of the relationship between models and metaphors. According to Bailer-Jones models provide access and interpretation of a phenomenon. Metaphors can provide this interpreted access but do not necessarily do so [1, p. 124].

I previously mentioned that my focus is primarily on analogue models. It is therefore best to discuss analogies themselves. What is an analogy? According to Bailer-Jones an analogy is "[..] often understood as pointing to a resemblance

between relations in the different domains, i.e. A is related to B like C is related to D" [1, p. 110]. I will later refer to this structure as A IS TO B LIKE C IS TO D [20]. Such structure is quite similar to Lakoff and Johnsons' description of the structure of a metaphor, namely X IS Y. Analogies within science have a similar function as metaphors. Bailer-Jones says that "[A]nalogies are employed in science to promote understanding of concepts. They do so by indicating similarities between these concepts and others that may be familiar or more readily grasped" [1, p. 110].

Here Brown's idea about metaphors as not having to make everything explicit is of importance. Metaphors contain within them analogies but they are implicit. If they would *have to be* made explicit they would probably not work. I myself should now also make clear that there are two ways in which analogies can be implicit within a metaphor: 1) implicit as implicitly understood, which is synonymous with intuitively understood/grasped, and 2) implicit as yet to be explored but already present within the structure. Metaphors implicitly imply analogies in that second sense. An analogue model is a model which consist of analogies. One can thus say that while a metaphorical concept at its base has the structure X IS Y it implies a deeper analogous structure.

I am uncertain how these two structures (the deeper structure and the analogous structure) are related. A possibility would be that X' IS TO X LIKE Y' IS TO Y. For instance, take the metaphor THE MIND IS A COMPUTER this has the deeper analogous structure: MEMORY IS TO THE MIND LIKE MEMORY IS TO A COMPUTER. Nevertheless, because models are metaphors and analogies are implicit within metaphors this also means that analogies are implicit within models.

What we learn from this is that models are indeed metaphors and that the models that I will focus on (analogue models) contain an implicit analogous structure. This analogous structure will later on make clear the process of theoretical postulation with respect to metaphors.

2.2 Metaphors as bridge between first-person perspective and third-person perspective

I would now like to show how the previous talk about models/metaphors is related to the first-person/third-person framework. In section 2.2.1 I will integrate metaphors in the first-person/third-person framework. This discussion of metaphors as bridge between first-person and third-person perspectives will

[20] 'IS TO' is short for 'IS RELATED TO'

be important in order to discuss how metaphors are also a source of confusion between those two perspectives (2.3)

I will then show how this extended framework is still neutral towards both first-person and third-person perspectives in section 2.2.2. In section 2.2.3 I will continue discussing this neutrality by focusing on the idea of metaphors as providing *partial* access.

2.2.1 Integrating metaphors within the first-person/third-person framework

In this section I will include the ideas that I discussed regarding metaphors within the third-person/first-person framework. In chapter 1 I have shown that third-person concepts are re-conceptualizations of first-person concepts. Metaphors make this relationship possible and function as a bridge between the third-person and first-person perspectives.

First-person concepts are based on having a particular perspective with perspective dependent experience. Such conceptualizations are formed by what appeared through such experience. Understanding of such a conceptualization of the world is done by finding correlations between first-person conceptualizations of things. Another important point was that the conceptualization of oneself was that of a person with person-like qualities (the importance of this should become more clear in chapter 3). Third-person conceptions are based on these first-person conceptions. The third-person point-of-view deviates from the first-person point-of-view by abstracting away from being about particular perspectives and their experiences. The third-person perspective is in that sense not restricted by direct observation because such direct observation is observation from a particular perspective. Such abstraction has two further ontological consequences for the third-person point-of-view 1) what is experienced to be there may not be there at all or in the form that it is experienced, and as a result 2) there could be *more* then what is directly observed. In this sense the third-person perspective is both an addition to and a subtraction from the first-person perspective. It removes what is particular to a self with a particular perspective (which is what makes persons what they are) and it adds the possibility of additional compositional elements (which were not observed in the first-person perspective). This de-compositional aspect of the third-person point-of-view does not only make possible further understanding of the correlations found within the first-person image, it also re-conceptualizes the things within the first-person perspective. In this section and the next section I will argue that metaphors are required for this process of re-conceptualization.

The abstraction that I talked about is an ontological abstraction, it is an abstraction of points-of-view not of their resulting images. As such knowledge of one's point-of-view (knowledge of one's ontology) determines but does not specify the resulting image. This is what metaphors do, they specify specific third-person conceptions. Metaphors make possible the transition from *image to image*[21].

To clarify let me first restate the basic definition of a metaphor: metaphors are the understanding of one thing in terms of another; structurally this would be: X IS Y. Y has to be familiar. In the case of a third-person/first-person re-conceptualization X is some first-person conception and is thus known. In order to comply to a third-person point-of-view its has to become abstracted (become decomposed and non-particular). This is done by drawing similarities between X and Y, where Y is some other conception which has a structure and/or a form which by virtue having such a structure/form both decomposes the source X and makes it non-particular[22].

For example, take the first-person conception of mind. Any first-person/third-person re-conceptualization of the mind would have the structure THE MIND IS Y. Within the field of AI one of such metaphors is THE MIND IS A COMPUTER. This effectively re-conceptualizes the mind in a third-person conception. To be clear, THE MIND IS A COMPUTER is the new third-person concept mind and 'THE MIND' refers to the first-person conception mind. Now we can easily see how third-person and first-person conceptions are related. They are related by metaphor and this becomes apparent when looking at the structure of such a metaphor. We can generalize this structure as: *a third-person conception is (ITS RELATED FIRST-PERSON CONCEPTION) IS Y*, where Y by virtue of having a certain structure/form has a decomposing and de-particularizing effect; Take our example of THE MIND IS A COMPUTER. In this example the source A COMPUTER has a de-particularizing abstracting effect; it removes elements from the concept of mind which are conceptualized as such because of the experience of a self with a particular perspective. For instance, one could have a first-person conception of the mind as non-material and not bound by deterministic laws. We know that a computer consists of only matter, therefore THE MIND IS A COMPUTER leads to an abstract concept of the mind mind which does not comply with the experience of mind - in other words, the effect is a de-particularization. Furthermore a computer consists of

[21] I am not saying that all use of metaphors result in a shift from first-person sub-image to third-person sub-image, just that first-person/third-person re-conceptualization of such sub-images is made possible by metaphors.

[22] I suppose the source could be either a first-person conception or a third person-conception. As long as it results in a third-person conception.

several components/parts non of which correspond to a first-person conception of the mind.

Third-person conceptions thus have to be metaphorical concepts in order for them to make sense. This because the theoretical postulation required for a third-person conception and its abstract form need to be understood in some form. This is exactly what metaphors do; metaphors are the understanding of in things in terms of that which is understood. That which is understood are first-person conceptions. I should add that not all metaphors are third-person conceptions and therefore not all metaphors function as a bridge between first-person and third-person perspectives.

In conclusion, metaphors function as a bridge because they can insert new theoretical postulation into an old first-person concepts as well as abstract away from such a conception; Metaphors function as a bridge because they are capable of making this new abstract third-person conception understandable, they are capable of providing a intermediary function which is also intuitively grasped.

2.2.2 Metaphors as bridge

Now that I have integrated metaphors in to the first-person/third-person framework I would like to elaborate on the position of neutrality that I am pursuing. I have given an account of metaphors as a bridge between first-person and third-person perspective, but the specific views from which I have borrowed the description of the function and nature of metaphor are perspective dependent. In order to give a neutral analysis of the conceptual confusion between first-person and third-person conceptions I will have to show how my own view can still be neutral considering this perspective dependency. I will start by talking about the difference between literal language and figurative (or metaphorical) language use and the difference between them. I will argue that what is considered to be literal and what is figurative is in part perspective dependent. This discussion will serve as a prelude into the broader view that the used description of the role of metaphors are perspective dependent.

According Lakoff and Johnson the use of metaphors is so widespread that virtually all conceptions are metaphors. But very few of these conceptions are recognized as metaphors. The expression of such unrecognized metaphors is considered by the general public to be literal language use. According to Bailer-Jones literal means: "[..] that an expression is not transferred from another domain" [1, p. 105]. This stands in direct opposition with what metaphors do. We can then in principle distinguish between literal language use and metaphor-

ical language use (both recognized and unrecognized). However, according to Bailer-Jones such a distinction is pointless because metaphorical language use is so pervasive. But one could still distinguish between certain degrees of metaphoricity[23] namely: new metaphors, a familiar metaphor that is still recognized as a metaphor, and such a familiar metaphor that it is (almost) not recognized as such.

These last types of metaphors are what Brown calls frozen metaphors [4]. According to Brown, frozen metaphors are those objects/facts that have become a part of the accepted paradigm; they are no longer seen as metaphorical, they are leftovers from active metaphorical use.

For example, we would not be surprised to hear someone say: "I need time to process this new information". We would not recognize that this expressions is entailed (and thus allowed and understood) by a metaphorical concept (i.e. THE MIND IS A COMPUTER)[24] [1, p. 115]. Our recognition of a metaphorical expression as a metaphorical expression is an indication of the acceptance of the involved metaphorical concept and of the extent to which such a metaphorical concept has shaped and impacted our conceptual framework.

This is somewhat odd because according to Brown: "*if taken literally, the metaphor must be patently absurd*" [4, p. 81]. For instance, sound is not the actual wave we see in a river. By making such a statement Brown is effectively saying that frozen metaphors are absurdities. Such a view leads to complications within the third-person/first-person framework. If we have a first-person/third-person re-conceptualization by means of metaphor then the resulting third-person metaphorical concept as a part of the third-person image (i.e. when it is a dead metaphor) would be (according to Brown) an absurdity. Within the third-person perspective it is not considered to be an absurdity. If all re-conceptualization are of a metaphorical kind then it can not be that (from a third-person perspective) metaphors always produce absurdities. Hence Brown's view on metaphors is perspective dependent.

Within a third-person perspective metaphorical concepts can be wrong but they can not be absurd, especially not necessarily absurd only because they are metaphors. A statement such as: "*metaphors are intended to be understood; They are category errors with a purpose, linguistic madness with a method*"[4, p. 82] implies that the third-perspective is in the end an insane enterprise. This is not how the third-person perspective sees itself. There may be token third-

[23]I borrowed this term from Bailer-Jones.

[24]To be fair it are not only computers that process things. Therefore, 'to process' used in a metaphorical expression could also be entailed by THE MIND IS A MACHINE. However, in this context the processing involves information, which is what computers do.

person images which are wrong but the ideal third-person image can neither be wrong nor absurd. Furthermore, if any (token) third-person image is deemed (as Brown implies) to be absurd because of its metaphorical structure then so must this ideal form. Therefore, if there is such a metaphorical structure and the ideal is not absurd (as it is necessarily seen from *a* third-person point-of-view) then a third-person image cannot be absurd (by virtue of having such a metaphorical structure).

Nevertheless, it is this recognition of this absurdity that according to Brown makes possible that metaphors work; Brown: "The logical, empirical or psychological absurdity of metaphor that has a specifically cognitive function: It makes us stop in our track and examine it." [4, p. 81]. Here too we have a statement that can not be accepted within the third-person perspective with respect to third-person metaphorical concepts. It should be clear now that Brown is describing a first-person perspective on metaphors.

Brown thinks of metaphors as only able to work when they are "consciously 'as if'"[4, p. 83]. Meaning that the new reality that metaphors create must be understood "from the inside as it were"[4, p. 85], thus "*as if* it literally were the case"[4, p. 85]. This seems to run contrary to the claim that metaphors are absurdities but what Brown means is that metaphors require "an attitude of 'as if' in which we suspend what is taken literally, and take as literal what we know to be absurd"[4, p. 85]". This last expression is very apt for *a first-person description of the third-person perspective*.

To emphasize this point, let us compare Brown's view on metaphors with the description given by Bailer-Jones of them. When Bailer-Jones says that metaphors provide access to a phenomenon she is implicitly saying that metaphors are not absurdities. To say that something is absurd is to say that something can not be taken seriously. If metaphors provide scientists with access to phenomena then they can not be absurdities. Bailer-Jones' description of models should thus be understood as a third-person view on metaphors.

I will maintain neutrality by not seeing metaphorical concepts (both the first-person kind and the third-person kind) as absurdities nor do I consider them to be truthful. With respect to their truth status I will merely see a metaphor-as-tool as making possible the re-conceptualization from first-person concept to metaphorical third-person concept.

2.2.3 Scientific perspectivism and partial access

Before going to the next section (which will be about the sources of conceptual confusion) I want to expand upon the idea of metaphors providing *partial* access. This section will thus function as further establishing my position as neutral and repeat what has been said before as preparation for the next section.

According to Bailer-Jones third-person description of models such a third-person metaphorical concept has provided *partial* access. This partiality of a token third-person conceptualization should not be confused with the incompleteness of the third-person image.

Metaphors only highlight certain aspects of the world. These metaphors are necessary for scientific progress. Metaphors can be used to provide further understanding of first-person conceptions of phenomena and their correlations. Such further understanding requires postulating unobservable theoretical entities. Two things are then needed: 1) the insertion of these unobservable entities (resulting in the decomposition of the first-person concept) and 2) that these new entities are understood (meaning that they are known prior to insertion) in order for the metaphor to be explored. If metaphors are used in such way that a re-conceptualization from first-person concept to third-person concept occurs then this third-person conception is abstracted away from its parent first-person conception in the sense that in no longer applies to a first-person perspective; it is no longer part of a particular perspective. If a metaphor breaks the original non-decomposability of a first-person concept then it has created a new third-person concept conception of that target.

Such a view that metaphors provide partial perspectives is compatible with Giere's position of scientific perspectivism. Such perspectivism should be distinguished from (objective) realism and (subjective) relativism. Objective realism can best be described as the view that there is one objective and *complete* description of the world which is both obtainable and is being made progress towards [16, p. 4]. Relativism would hold that there are only perspectives on the world and that none are or can be any better than any other [16, p. 13]. Perspectivism is the position that one views the world from a certain perspective (which is partially constructed) and that within that perspective things can be correct or incorrect. People holding the same or similar perspectives would then have a similar view on what is objectively true. A perspective that does not hold on to its own principles is worse then one that does. For example, an incoherent perspective is worse then a coherent one (assuming both adhere to the principle of coherency).

What we can learn from this that it is indeed still appropriate to use Bailer-Jones' description of models as providing partial because within a third-person perspective it can still be acceptable to see them as such.

2.3 Sources of conceptual confusion

In this section I will show why the use and nature of metaphors (which is a bridge between the first-person and third-person perspective) is responsible for (possible) conceptual confusion between third-person conceptions and first-person conceptions.

I will discuss three related sources of conceptual confusions. In order of importance they are: 1) polysemy 2) frozen metaphors 3) the idea that third-person conceptions are *mere* metaphors. They are related because they are all about metaphors. One can thus observe prior to providing details about these sources of conceptual confusion that what makes possible re-conceptualization between the first-person domain and third-person domain (i.e. metaphors) is what causes conceptual confusion with respect to these domains.

2.3.1 Polysemy

Even though metaphors are crucial to the (third-person) scientific enterprise they may also create confusion. Metaphors create new conceptions of first-person conceptions; one might think that it is the old conception that is being discussed when discussing this new conception. I have talked about how these first-person concepts can be the basis of new third-person concepts. I have not yet said anything about the terms denoting those concepts. I will do so now. Doing so should provide insight in to why there is conceptual confusion.

While the conception of something shifts from first-person to third-person during abstraction and theoretical postulation (i.e. re-conceptualization), terms may not. Terms cannot be abstracted but they can change. A term A can become A'. It is possible that after metaphorical abstraction a new term will be used for this new conception. So, term A may refer to first-person conception X and after metaphorical re-conceptualization of X we have a term A' referring to this new third-person conception conception X'. However, this does not have to happen. After re-conceptualization we could have one term denoting two concepts. We could end up with an ambiguous term. In the context of this thesis such ambiguity is between first-person and third-person conceptions. Such a term with multiple *related* meanings is called a polysemous term. Taylor

contrasts such polysemous terms with monosemous terms: "A monosemous lexical item has a single tense"[38, p. 99] whereas "[a] polysemy is the association of two or more related senses with a single linguistic form"[38, p. 99]. This is the case with the term 'mind'.

For example, take the third-person metaphorical concept THE MIND IS A COMPUTER. The term 'mind' can refer to THE MIND IS A COMPUTER and to the first-person conception of mind. This last conception of mind is the 'THE MIND' part of the metaphorical concept THE MIND IS A COMPUTER. The phenomenon mind can be viewed from two different perspectives each having a different yet related conception of the mind, but both being referred to with the term 'mind'.

We can thus conclude that one reason why conceptual confusion arises is because of the occurrence of polysemous terms whose senses are distributed across both first-person perspective and third-person perspective; it then becomes easy for someone to interpret a term as referring to a third-person (or first-person) conception while it was intended to refer to a first-person (or third-person) conception (or vice versa).

2.3.2 Frozen metaphors

I have already described that metaphorical concepts are not necessarily recognized as such. These frozen metaphors are another reason for conceptual confusion. This reason is thus strongly related to polysemy as a reason for conceptual confusion. If a metaphor is not recognized as such this has the effect one can no longer recognize its term as a polysemous term. If that is the case then one might discuss such a metaphor as if no one would be able to interpret it differently. If a term is polysemous it may very well be interpreted in this other way.

2.3.3 A third-person conception as merely metaphorical

Another reason for confusion is the idea that some description of a phenomenon is merely metaphorical. In the case of the mind we find Eliasmith arguing we should move beyond the use of metaphors [13]. In short, such a claim is irrelevant. First off, almost all our conceptions are metaphorical. Second, models are needed to describe theoretical postulation and those models are metaphors, therefore science is built on the use of metaphor; postulation can not do without. Therefore if we want to understand the mind besides pre-conceptual

experience, it will always be with the use of a metaphor. The statement that "THE MIND IS X is just a metaphor" is meaningless.

For example, one could roughly sketch three different metaphorical re-conceptualization of the mind in order to understand it: 1) symbolism, which uses the mind as a computer metaphor, 2) connectionism, which considers the functioning of the mind to be like the functioning of the brain, and 3) dynamicism, in which the mind is a Watt Governor [13, p .493-494]. These three approaches follow a strategy of not analyzing the brain-mind correlation directly, by analyzing the brain itself but by making theoretical idealizations (in forms of postulations) of the mind. This is done by positing a mind and exercising a priori theorizing on its features and properties [7, p. 266]. These alterations of the first-person conception of the mind consist of postulating a mind that is decomposed and thus no longer a part of the manifest image. Such metaphors are not just used for explanatory purposes but also for a clear conceptual understanding of the metaphorical target (the mind).

One could also argue that the specific metaphor used is not an appropriate metaphor. It is considered to not be appropriate because it does not fit with a certain idea of the world. Often times what is used to disqualify a metaphor as appropriate in the philosophical discourse about the mind is the lack of intuitive appeal. If some metaphor does not have metaphor has no intuitive appeal it is considered to be incorrect. This could be valid but it does not disqualify the use of metaphor nor does it make the claim that some description is merely metaphorical any more meaningful.

One reason why one would suggest that some conception is merely metaphorical is because of the occurrence of frozen metaphors. If one does not recognize that many conceptions are metaphors (which is possible because a lot of conceptions are frozen metaphors) then it becomes easy to say that some particular conception is a mere metaphor. *We can thus see that all three sources are related because of the nature and structure of metaphors; Metaphors thus function as a bridge between first-person conceptions and third-person conceptions but are also a source of confusion.*

2.4 Conclusion of chapter 2

In this chapter I have described metaphors as a bridge between third-person and first-person conceptions. I have focused on the ability of metaphors (as tools) to both abstract and decompose first-person conception. The resulting third-person conception can be confused by a first-person conception because

of this metaphorical relationship between the two. This can happen because the same term is used to denote both first-person concept and third-person concept, because the third-person concept is a frozen metaphor or because such a metaphorical concept is seen as merely metaphorical. In the next chapter I will explore the resulting conceptual confusion even further, In that chapter I will primarily focus on the polysemous nature of certain terms in particular 'understanding', 'ability' and 'know-how'.

3 Thought-experiments: Two case studies

In this chapter I will describe two thought-experiments (the Chinese room (3.2) and Mary's room (3.3)) and show how the first-person perspective and the third-person perspective produce different views on the validity of those thought-experiments. Both of these perspectives make different claims about them. These two distinct perspectives on the mind have two distinct conceptions about the mind. Having different conceptions about the mind correspond to having different intuitions about the mind; thought-experiments pump up these intuitions (even unintentionally) and cause their corresponding replies. The diversity of intuitions, and the conceptual confusion that arises (because of reasons given in section 2.3) will be analyzed using the framework that was set up in chapter 1 and chapter 2. Before giving such an analysis I will first provide a broad overview of what thought-experiments are.

3.1 Thought-experiments

In this section I will provide a brief overview of what thought-experiments are (3.1.1), what their function is (3.1.2), and why I use thought-experiments (3.1.3) as case-studies.

3.1.1 What are thought-experiments?

Philosophical thought-experiments about the mind test certain conceptual claims. These philosophical thought-experiments should not be confused with scientific thought-experiments. This thesis only deals with philosophical thought-experiments.

There are different opinions about whether or not philosophical thought-experiments are capable of proving some idea. I will look at thought-experiments as being capable of convincing someone of a position because they are argumentative. This does not mean that I endorse thought-experiments as a *proper* tool to determine whether something is true or false.

Nevertheless such argumentative thought-experiments should be contrasted with neutral thought-experiments. In the neutral cases they simply show-case an idea without trying to show that this idea is correct (e.g. swampman[6]). In the argumentative cases they try to convince the reader of a certain position. My focus will be on such *argumentative philosophical thought-experiments* and henceforth 'thought-experiment' should be read as such.

3.1.2 What is the function of a thought-experiment?

I will speak about 'thought-experiments' as if they can provide insight into particular philosophical problems. This view on thought-experiments is not shared by all. For instance, Wilkes describes thought-experiments as primarily misleading, and would prefer to use real case examples in order to get more conceptual clarity [41][25]. (She also makes the important distinction between thought-experiments that are used to test the truth of a thesis and to test the scope/range of a concept.) The reason behind this is that for an experiment to work, it is required that there are certain known constraints on what can possibly have an influence on the experiment so that we know what aspect leads to the outcome of the experiment [41, p. 7]. In a philosophical thought-experiment we do not know which aspects of our imaginary world have remained unchanged and which have not *with respect to the real world*. If we do not know what has changed in this new imagined setting we do not know what the consequences are of these unknown changes. In other words, thought-experiments have ambiguous contexts which lead to ambiguous results [41, p.8].

There is also a second problem, namely the uncertainty regarding the use of our intuitions and imagination. Intuitions about a thought-experiment can very well contradict each other, which cast doubt upon the correctness of such intuitions [41, p. 16-17]. Furthermore, that we can imagine something does not mean that the thought-experiment is true or meaningful. Further investigation of a thought-experiment often shows that certain things that are assumed to be possible are actually impossible.

Wilkes may be correct about the impractical use of philosophical thought-experiments. But my concern with philosophical thought-experiments is not with respect to their correctness (or practical usefulness). The purpose of my analysis of thought-experiments is to show that there is a conceptual confusion between conceptions belonging to the first-person image and the third-person image. My analysis of though-experiments will highlight this confusion. As such, thought-experiments serve a purpose within this paper.

Such use of thought-experiments can be complemented with Dennett's view of thought-experiments as intuition pumps[8, p. 12][26]. According to Dennett thought-experiments provide a setting in which a certain intuition (or intuitions)

[25] Although she considers some to be not misleading.

[26] To be fair, I am uncertain if Dennett considers all *philosophical* thought-experiments to be intuition pumps. However, the examples that he offers of 'correct' thought-experiments are all non-philosophical whereas the 'misleading' thought-experiments are all philosophical ones.

is pumped up which is used as as a dictate in favor of a certain idea[27]; Dennett: "[intuition pumps] entrain a family of imaginative reflections in the reader that ultimately yields not a formal conclusion but a dictate of 'intuition'"[8, p. 12]. This is why it is more correct to say that thought-experiments do not prove but rather convince. That does not mean that philosophical thought-experiments are useless. They can provide an easy overview of the problem. They do so by creating discussion about a specific subject illustrated with a specific case study. By focusing on a particular intuition pump people can efficiently talk about a concrete and specific example.

This is not exactly the type of usefulness that I am concerned with. What will be useful in the context of this paper are intuitions as perspective dependent intuitions. What I will do is study thought-experiments as intuition pumps which pump perspective dependent intuitions. I will elaborate on this in the next section.

3.1.3 Why use thought-experiments?

According to Watt, there always is tension between the logic of a model (i.e. to what extent such a model makes logical sense) and its intuitive appeal (i.e. the extent to which we intuitively agree with this model). Therefore a source of disagreement about the mind centers on the conflict between the intuitive appeal and the logic of the metaphors used regarding the mind [40, p. 51]. This conflict is often pointed out by thought-experiments. Looking ahead, in the case of the Chinese Room (an explication/variant of the information processing metaphor [40, p. 55]) we (or at least some) fail to identify with the room (as is required by the system reply, which basically states that the Chinese room as a whole understands Chinese); we fail in anthropomorphizing the room [40, p. 54-55]. Such a reply might be a logically correct reply but it does not have intuitive appeal.

Such identification plays such a crucial role because those who study the mind have minds themselves; we know what it is that we are studying; this *first-person knowledge* of the mind has an effect on our perception of minds and certain proposed metaphors of the mind [40].

If a thought-experiment is about the mind then some entity *which is a person* is (supposed to be) included in it, because only *persons* have minds (according to our current first-person perspective)[28] . Therefore any imagination of the situation includes imagining such an entity.

[27] Unlike, for example, Galileo's leaning tower of Pisa thought-experiment.
[28] Which may include animals.

According to Schoemaker there are two ways in which such an entity can be imagined: from a first-person point-of-view and a third-person point-of-view. These terms do not denote what I mean with 'first-person point-of-view' and 'third-person point-of-view' but are similar. What he means with those terms is that one can imagine being the entity (i.e. first-person point-of-view) or imagine that there is some entity (i.e. third-person point-of-view) [31, p. 7]. I would like to add that a thought-experiment that asks for a first-person perspective on the entity (i.e. imagining being that entity) is reminding us of our first-person conception of the mind and asking us to find the similarities (and dissimilarities) between this conception of the mind and the entity. If such a thought-experiment is about rejecting some specific third-person conception of the mind we will always find dissimilarities when asked for a first-person perspective, because we are asked to make a comparison between a third-person conception and a first person conception which have their own (though related) conceptual domains.

In the case of imagining a subject this requires a comparison with that subject. The likelihood of success of such comparative identification depends on the similarity between our ordinary experience of mind (which is how we intuitively perceive the mind; the first-person conception of mind) and the system itself. According to Watt, if such an identification finds resistance, several forms of restructuring or reinterpretation of the system are attempted. These attempts themselves will then go through the same process [40]. Normalization on such a system is the result of a failure of initial identification; its restructuring should, in the ideal case, lead to a modified system which is easier to identify with and thus has a stronger intuitive appeal [40, p. 60].

Watt's view can also be phrased as the clash of the two perspectives as the result of our attempt to provide a third-person conception of the mind. The first-person perspective (or rather subjective tools, identification and intuition) plays a role in determining the fitness of a model/metaphor.

If one is committed to understanding the world from a third-person perspective, one is also committed to some metaphor that abstracts away from such a first-person perspective. What often happens is that the reader does not comply with the intention of the writer of taking a first-person perspective resulting in a third-person perspective counter-response (i.e. reinterpretation or restructuring) to such thought-experiments. This is where conceptual confusion becomes most apparent. This is the reason why thought-experiments are useful case-studies to investigate conceptual confusion. I will discuss two of these thought-experiments (the Chinese room and Mary's room) in depth (3.2 and 3.3).

In summary, thought-experiments pump up intuitions. These intuitions depend on the perspective taken. Although the writer of a thought-experiment may have intended (as well as given the form of the thought-experiment attractive force towards that perspective) the reader to take a first-person perspective this may not always happen. Therefore, different intuitions can be instilled. If this happens without being aware of the fact that these intuitions are perspective dependent then conceptual confusion can arise. I will now show that this the case by first discussing the Chinese room (section 3.2) and then Mary's room (section 3.3).

3.2 The Chinese room

In this section I will discuss the Chinese room and how it relates to the framework that I have described. I will do so by first giving a reiteration of the original Chinese room thought-experiment (3.2.1) and a brief overview of the anticipated replies to it (3.2.2). Then, to understand Searle's intention in giving this thought-experiment, a closer examination of Searle's view on reality and mental phenomena will be given (3.2.3/4). Following this, it will become clear why Searle will never be satisfied with the kind of answers given (3.2.5) because Searle takes a first-person perspective and the replies are of a third-person kind. Subsequently, I will show why the replies have the specific third-person content that they have (3.2.6). I will then give a general account of what both first-person and third-person conceptions of understanding are and their relationship (3.2.7/8). This should highlight the conceptual confusion even further (3.2.9). I will finish with a discussion of Searle's own idea of the mind with respect to the Chinese room after which it will be clear that he too suffers from conceptual confusion(3.2.11).

3.2.1 The original Chinese room thought-experiment

Searle gives the following thought-experiment to show that strong AI is wrong [32, p. 419]. *Imagine* that *you* are locked in a room. You do not understand Chinese. You are given in chronological order: a large batch of Chinese writing; a second batch of Chinese writing and a rulebook written in English; and a third batch of Chinese writing and some English instructions. The people who are giving you these Chinese writings call those batches a script, a story and questions respectively.

The rulebook is used to correlate the second batch with the first batch. The instructions are used to correlate the third batch with the previous two batches.

These instructions also tell you how to give back certain Chinese symbols (in response to symbols given in the third batch). The external viewers (outside of the room) call these responses answers to the questions. The rules given are called a program (by those viewers)[29].

Searle concludes from this experiment that the external viewer would initially attribute understanding to the room purely based on observable behavior. However, this does not mean that you understand Chinese. What you have been doing is manipulating formal symbols. And this does not entail any form of understanding. While you are imagining yourself correlating formal symbols you do not at any point see yourself as understanding Chinese purely on the basis of manipulating those formal symbols. Thus, Searle concludes, to understand (Chinese or anything for that matter), one needs something else than formal symbol manipulation [32, p. 419].

3.2.2 The four anticipated responses

Searle anticipated four distinct replies on his Chinese room thought-experiment: the systems reply where you as an individual do not understand Chinese, but the system as a whole does [32, p. 421]; the robot reply where you are replaced by a computer - this computer would then interact with the world beyond mere linguistic communication (e.g. see, hear, smell etc.); in other words, create a more interconnected situation [32, p. 423]. The brain simulator reply where the program is replaced with a simulation of the firing of synapses of a native Chinese speaker [32, p. 424]; the combination reply which combines all three replies together into one solution [32, p. 424]. All these replies have actually been both given and supported by scientists and philosophers. Searle is not satisfied with any of these possible replies. To see why, we would have to elaborate on Searle's view of reality and mental phenomena[30].

3.2.3 What is Searle's view on reality?

In short, Searle states that "*not all of reality is objective*"[33, p. 19]. Mental states are subjective and real. They are a subjective part of reality. We can give an objective account of beliefs, minds etcetera in order to attribute them, but

[29]This thought-experiment has been rephrased in a simpler form. In its simplest form there is only a rulebook which states for every input (i.e. questions) which output to return (i.e. answers).

[30]To do this I refer heavily to Searle's book: "The rediscovery of the mind" published in 1992, 12 years after the creation of Searle's original Chinese room experiment. I am assuming that 1992 Searle has very similar ideas about the mind as 1980 Searle.

ontologically mental states are first-person phenomena [33, p. 16]. Therefore, ontologically speaking not all of reality is objective [33, p. 19].

I should note that Searle uses the terms 'objective'/'subjective' differently but still similarly as Nagel's and likewise with 'third-person'/'first-person' as compared to my designation of those terms. What Searle means by subjectivity is an "ontological category"[33, p. 94] which existence is a "first-person existence" [33, p. 94]. This first-person existence implies that it is only (introspectively) accessible to oneself. Objectivity on the other hand is that which is "equally accessible to any observer"[33, p. 16] meaning no particular observer. One example of such equal accessibility would be "the objectivity of external behavior"[33, p. 16]. Searle's use of 'objectivity' does not restrict itself to decomposable phenomena but only to that which can be observed by no one in particular (cf. Nagel). My description of the third-person perspective is thus somewhat more exclusive. Searle's view of a third-person point-of-view as simply an abstraction away from introspection is enough to include observable behavior as an objective part of reality. I would see this as still a part of the first-person perspective because even though it has no subjective qualities it does not violate those presupposed qualities either.

It is only at the point when this behavior is equated with the mind that one has a third-person perspective on the mind because its first-person features are then abstracted away. This last consequence is why Searle is against the understanding of the mind from such a third-person point-of-view.

What we can conclude is that it is appropriate to see his use of the terms like 'objective','subjective','third-person' and 'first-person' as comparable to my use of these terms (within the context of this paper). Even though his use of the word 'ontological' has a different connotation compared to my first-person/third-person framework; when Searle states that the world is partially ontologically subjective he means that existence is partially subjective whereas my description of the first-perspective as conceiving the world in a certain ontological manner merely means that within this perspective the world is conceptualized as such (without making any truth claims regarding its existence). It is thus also appropriate to conclude from Searle's statement that "not all of reality is objective" that such a position entails a first-person point-of-view[31]. This kind of reasoning will also apply to everything else I have to say about Searle.

[31]This is not exactly correct because within the first-person perspective all that exists are first-person conceptions and therefore within that perspective *all* of reality is subjective. Which is not Searle's position.

3.2.4 What is a mental phenomenon according to Searle?

Searle sees mental phenomena as higher-level biological properties of neurophysiological systems. Additionally, those biological properties are irreducible [33, p. 28]. The reason for this is that ontologically, these mental states are a first-person ontology [33, p. 16]. In other words, the essence of mental content is subjective and therefore it can not be naturalized, hence it is irreducible [33, p. 50].

Functionalism, according to Searle, is constructed with the aim of attributing intentional states (such as understanding) based on third-person perspective evidence. Actual mental states are not merely attributed, they exist [33, p. 20]. It is not just function that is irrelevant to their existence; (observed) behavior is so as well (e.g. attribution of mental states due to their observed corresponding behavior is not what is responsible for their existence). Behavior, functional role and causal relations are simply not relevant to the existence of mental phenomena (in the ontological sense) [33, p. 69].

Having this in mind Searle can easily say that beliefs and desires are not postulated to exist rather they exist by virtue of being experienced [33, p. 59]. The subjective nature of mental states is what makes them real. This is then what would be essential to mental states, their subjective properties. To put it briefly; a mental phenomenon is (essentially) subjective.

3.2.5 Why is Searle not satisfied with the answers given?

Now we can see why Searle would never be satisfied with the kind of answers given. All answers are third-person answers. A third-person perspective only takes into account the objective. A third-person perspective can only give third-person answers. It can only give objective answers. Functionalism for example can only describe that which is accessible to all (e.g. from a third-person perspective) and therefore can never truly say anything about mental phenomena, which is only accessible to one (e.g. subjective). Summarized, Searle will never be happy with a third-person answer, since it is not a first-person answer.

3.2.6 Then why do the replies have the third-person content that they have?

We can now see where the robot reply and the systems reply go wrong. Both simply do not take into account the first-person perspective. Their description

(for instance in the form of functionalism) does not take into account that all mental states are initially known by experience.

For example, when the systems reply responds to Searle by stating that we should see the system as a whole as understanding Chinese, one misunderstands Searle's understanding of understanding. Although Searle never defines understanding as such, he does implicitly describe understanding as a mental state and it is thus, for him, a subjective state (which is a first-person conception)[32]. Any third-person description of the Chinese room will fail to insert this subjectivity into its description.

Searle argues that people who would reply in such manner are under the spell of some kind of dogmatic ideology or religion. I am not convinced that this is the whole truth (if there is any truth to it). Those scientists and philosophers are simply responding to the claim that the person in the Chinese room does not understand *from a third-person perspective*, and that the entity (or something else about the room) should be able to understand Chinese from that perspective[33]. Someone taking a third-person perspective on the program itself may very well speak of the understanding of Chinese, depending on the metaphor used to model the mind and the interpretation of what understanding would mean within that model.

However, Searle was not asking for such a view. He was already anticipating such responses (i.e. third-person responses) as wrong responses (for him). The right kind of replies are not the kind of replies his third-person opponents can give. When they hear/read 'understanding' it is interpreted as a third-person term. This creates some very different intuitions about the Chinese room and therefore also a distinctive kind of reply.

The robot reply makes a similar mistake. Having the Chinese room interact more with the world, that is, have more computational interaction with the world, does not make it acceptable by the first-person perspective. A computational description does not show that the Chinese room understands anything in the first-person sense of the word. A description in third-person terms does not force us to accept the claim that it understands in the first-person sense of the word, because a first-person conception of understanding is different from a third-person conception of understanding.

[32]This does not mean that mental states are necessarily *only* conceptualized from a first-person perspective, but how Searle has conceptualized mental states is.

[33]Not just when judged by an external observer (standing outside the room) but also when looking inside the room at its decomposed mechanisms.

3.2.7 What is the first-person conception of understanding?

According to Jaynes, understanding is the feeling of familiarity [20]. This is a first-person conception of understanding. It corresponds with what understanding introspectively appears to be. When pondering long about some difficult text, we eventually come to understand this text. There is than a sensation of understanding, usually without knowing what was required to understand the text. Which could be included in a third-person conception of understanding. Such a first-person conception of understanding is a very basic - it can not be decomposed without becoming a third-person conception. We just have a sensation of understanding and that is all. Searle is right; understanding in this sense (i.e. from a first-person perspective) is subjective and irreducible.

3.2.8 What is the relationship between the two conceptions of understanding?

To see what understanding in the third-person perspective is, one should first realize that although we are often only aware of the sensation of understanding we would be able to to detect patterns in our behavior leading to such a state. We can draw correlations between our observable behavior and the sensation of understanding. In procedural terms we are able to see what leads to that sensation of understanding from which we derive a first-person conception of understanding. Such a procedure can be further decomposed but it would be devoid of any subjectivity, that is to say such a conception would be formed by taking a non-particular point-of-view.

Such a third-person conception of understanding is an abstraction away from its derived first-person conception. Searle himself seems to allude to something like this when he says that: "subjectivity is removed from the mental by rewriting ontology in terms of epistemology and causation" [33, p. 21]. This conception of understanding is not specific to the experience of an individual, which is more abstract as it now has a broader application, it is now capable of being part of all minds. One example of such a third-person conception of understanding would be the correct manipulation of symbols. Such a conception is based on the third-person metaphorical concept THE MIND IS A COMPUTER.

Searle and others may argue that such a metaphorical concept is just a metaphor. This claim is missing an important point. Models are metaphors and are necessary for science to work. Therefore Searle misses an essential point about the use of metaphor and its consequences. By using metaphor we are no longer talking about the same thing when we are discussing third-person and first-person

conceptions of the mind. Additionally, asking people (specifically scientists and philosophers) who take a third-person perspective (specifically a computational one) to take a look at the Chinese room produces different answers (in the form of objections and modifications of the Chinese room).

This may be interpreted as them having trouble figuring out how the Chinese room achieves any first-person understanding. However this may not be the case; in some cases it is an attempt to clarify what a third-person conception of understanding is, not a first-person one. I will address these issues in the next section.

3.2.9 Why do we get any replies at all?

If one takes a third-person point-of-view on understanding then one would have much less difficulty accepting that the person in the room (which is you) understands Chinese. An abstract third-person notion of understanding leads (or could lead) to a description in the form of a program understanding Chinese. Therefore, you do indeed understand Chinese when running a program. Yet we do get replies in the form of objections/modifications (for over 30 years); these objections/replies show that something must go wrong regarding the discussion revolving around Chinese room (and the third-person conception of understanding).

Two possible answers: 1) the third-person conception of understanding is still not properly defined; it is in its infancy and these replies are a way of working out what the specifics of this particular abstraction would be 2) there is confusion about the proper domain of the two notions of understanding; there is an attempt to provide a first-person conception of understanding by taking a third-person perspective (i.e. conceptual confusion).

The robot reply falls under possibility 2). This reply attempts to make it easier to identify with the Chinese room. In the original thought-experiment we are asked to imagine being the person in the room. Searle's response to the robot-reply is to also have the man in the room receive (Chinese) input symbols corresponding to other sensory information. By making the room (analogous to the robot) more human-like we would expect it to have a more human-like experience. Our usual experience of (and interaction with) the world is more than communication; when we pretend to be the person in the Chinese room we have some difficulty seeing how it could understand because we experience the world to be more then just consisting of linguistic communication (e.g. hear, see, smell etc.). If the imagined entity is more like us (as conceptualized from

a first-person perspective) by having a more diverse sensory input it becomes easier to identify with it[34]. But this adaption of the Chinese room is still one in which the mind is pictured as a symbol manipulator; this conception is a third-person conception and therefore is no longer about a first-person conception of understanding (the same applies to the robot itself).

Now for the systems reply: This reply falls under possibility 1. The systems reply indicates an uncertainty within the third-person perspective about what is doing the understanding. If we would know what a third-person conception of understanding implies then we would not have this problem. Then it would be clear whether it is the individual that has understood Chinese, the system as a whole or for that matter the rulebook itself. Another reason for thinking that the system approach is taking approach 1) is that it seems difficult to see how it could be a form of approach 2). Certainly imagining being a room with a person in it (holding some written texts) is even more difficult then imagining being the person in the room!

Another indication that this is an attempt at clarifying the third-person conception of understanding is what Searle calls its absurd consequences [32, p. 423]. According to Searle accepting the systems reply has the implication that a whole range of systems are capable of understanding. We can not accept this for a first-person conception of understanding because only people can understand things. I would argue that we could accept this for a *third-person* conception of understanding. It would be a matter of creating distance between the two notions of understanding, which depends on the degree of abstraction (away from a first-person conception of understanding), which increases the range of applicability of 'understanding'. It seems unlikely that we would ever take this route considering the strong urge to keep the first-person and third-person conceptions as close together as possible.

3.2.10 But what about the brain simulation reply?

I have purposefully not discussed the brain simulation reply. Mainly because (as Searle has noted) it is not a reply that talks about the computational mind, thus it does not save Strong AI and is not really directed at the Chinese room as an argumentation against Strong AI. It is nevertheless possible to study this reply in relationship with the third-person/first-person framework.

The brain simulation falls under approach 1). It is an attempt to provide a first-

[34]For reasons of convenience I am excluding deaf and/or blind people, but it would be interesting to see if these people have distinct intuitions about this reply compared to people who are not impaired.

person conception of understanding from a third-person perspective. Although we know we have a brain and from a first-person perspective we know that this brain is correlated to our (first-person conception of the) mind, this does not mean that the brain as a *neurophysiological system* can be equated to the first-person conception of the mind. The brain as a neurophysiological system is a third-person conception of the brain because we can not directly observe neurons, it is a decomposed system. To say that the mind operates as neurophysiological brain is a re-conceptualization of the mind into a third-person conception. Such a view on the mind can lead to a third-person conception of understanding but such a conception would no longer corresponds to the first-person conception of understanding.

This can be clarified by looking at Searle's reformulation of this reply into an adapted version of the Chinese room. The main reason given why a simulation of the firing of neurons would produce understanding is because such simulation would simulate parallel processes. Searle proposes that instead of seeing the man in the room returning Chinese symbols as stipulated in the rulebook he now turns on and off the valves of a water pipe system as stipulated by some other rulebook. Each water connection is supposed to correspond to the synapse of a brain. And the right configuration of these water connections returns the appropriate Chinese symbol. Searle then asks where in the system the understanding of Chinese occurs. There is nothing in this adaption that leads to an identification with this system as person and therefore it can not understand Chinese (as seen from a first-person perspective).

3.2.11 And what about Searle's solution?

Searle's solution is to simply eliminate the computational mind (at least as independent of the brain) on the basis that the mind and its mental domain are subjective and the computational mind is not and therefore is not a mind at all. What Searle is left with is an account of the relationship between neurophysiological brain and first-person conception of the mind. Searle's solution is in this sense no better. He has simply demonstrated a need for a first-person account of mind, which he in turn wrongly presupposes can be given by a third-person account of the mind, one that is not in the form of computationalism/functionalism.

Both Searle and his computational opponents are primarily talking within their own conceptual domains failing to realize the distinction and relationships (in terms of abstraction and postulation) between the concepts within those domains. This leads to confusion. Repliers do not see that their computational

concepts do not apply in a first-person domain and Searle does not see that his first-person concepts do not apply in the third-person computational domain. Searle simply rejects the third-person computational domain and therefore does not have to justify that particular step. Repliers do not (necessarily) reject the first-person domain and simply remain confused, either about the third-person computational notions themselves (at least those derived from subjective experience) or by somehow thinking that computational notions require (and thus can have) something more identifiable (due to the history of those terms, being derived from first-person conceptions).

3.2.12 The Chinese Room as a case of conceptual confusion

We have seen that there are (broadly speaking) two distinct conception of understanding. These conceptions are used and understood as if they are one conception; as if 'understanding' has only one sense and is thus not a polysemious term. This leads to replies aimed at rejecting the Chinese room argumentation like the system reply and the robot reply without being about a first-person conception of understanding which was what the Chinese room was originally about. Searle himself does not recognize this confusion either and thus participates in it by rejecting these replies as incorrect (while they are at best misplaced). In conclusion, the Chinese room highlights the conflict between first-person and third-person perspective as well as the conceptual confusion that can arise due to polysemous terms whose senses are distributed across both those perspective.

3.3 Mary's room

In this section I will discuss Mary's room and how it fits with the framework that I have set up. I will start by reiterating this philosophical thought-experiment (3.3.1) and then show that it is based on one intuition and one important inference (3.3.2), the knowledge intuition (the idea that Mary learns something new) and the anti-physicalist inference (the inference that if Mary learns something new then that means that physicalism is false).

I will then describe Mary's room as a thought-experiment that is about the relationship between third-person perspective and first-perspective and that the positions regarding this relationship are perspective depended (3.3.3). I will do this in part by showing that the knowledge intuition (KI) is only acceptable within a first-person perspective. It is in this respect that the analysis of Mary's room will differ from the Chinese room. The Chinese room showed cases of conceptual confusion in a context where the two perspectives were present but

not explicitly discussed. Mary's room on the other hand is explicitly about these two perspectives. The Chinese room showed conceptual confusion because of not being aware of the presence of these two perspective. Mary's room on the other hand will show what happens when the (deeper) structure and relationships of these two perspectives is neglected. A different (complementary) analysis is thus required.

Following this argumentation I will show that such a relationship (and consequently Mary's room) looks like from within these perspectives (3.3.4), this should make clear the mentioned dependency. Of course people do not necessarily take only one of these perspectives but even when they do not their position may still depend on their perspective. This will be discussed in section 3.3.5.

In section 3.3.6 I will discuss the rejection of the premise that we know what it means for Mary to know everything about colour, thereby rejecting the knowledge intuition and section 3.3.7 will show a basic outline of accepting this knowledge intuition but rejecting the anti-physicalist inference. Section 3.3.8/9/10/11 will be about one of such counter-argumentation, namely the ability hypothesis. Section 3.3.8 will reiterate this reply and section 3.3.9 will show the conceptual confusion on which this argumentation is based by using the previously constructed third-person/first-person framework. In section 3.3.10 I will speculate why such confusion may occur.

I will then compare this ability hypothesis to a similar reply namely the acquaintance hypothesis (3.3.12). I will show that a corresponding analysis applies to that acquaintance hypothesis (3.3.13).

3.3.1 The original Mary's room thought-experiment

The original Mary's room thought-experiment as constructed by Jackson[18, p. 130][19, p. 291] goes as follows:
Mary is in a black and white room. She observes the outside world by viewing a black and white television screen. Everything she sees while remaining in her room is in black and white. Additionally, she learns all there is to know about the neurophysiology of (colour-) vision, what Jackson calls "physical information"[18, p. 127]) which is all the information provided by the physical, chemical and biological sciences[35]. She knows all the physical information provided by the complete scientific description of colour-vision. After she has learned all that information she is placed outside the room where she experiences colour, the colour red. According to Jackson Mary would now know more than

[35]This also includes descriptions of functional mental states

she knew before leaving her room; "she will learn something about the world and our visual experience of it."[18, p. 130]. Jackson does not specify what this 'something' is apart from stating that she learns something about the mental life of others[19, p. 292]. Whatever the case Mary learns something new. Therefore physicalism (the position that "all information is physical information"[18, p. 127]) is false. This (kind of) argument against physicalism is also known as the knowledge argument [22, p. 5].

3.3.2 What is Jacksons' knowledge argument based on?

Jacksons' view (until he changed his position) is a straightforward denial of physicalism. Physicalism is the idea that all information is physical information. According to Jackson there is also information that is not physical information; our experience of the world gives us this kind of information. Phrased slightly differently, not everything in the world can be described in physical terms. According to Jackson qualia is one (and possibly the only one) of those things not describable in physical terms.

There is an intuition that underlies this idea. This intuition has also been named the knowledge intuition by Stoljar and Nagasawa[22]. This knowledge intuition can be expressed as follows: *knowledge of a physical kind (or any sort other than phenomenological knowledge for that matter) will never produce knowledge of a phenomenological kind* [22, p. 2-3]. This intuition in itself does not entail that one can not describe the causes and effects and/or function of the experience in physical terms (although Jackson has a different view, he sees qualia as epiphenomena [18, p. 134]). It just means that a physical description of an experience (or anything relevantly related to it) does not provide the information contained within that experience.

The knowledge intuition (KI) is a part of the argument but may itself not be considered to be correct. It is an intuition that is pumped up by the story that Jackson tells. When it is pumped up it becomes a part of the argument. This intuition by itself does not imply falsity of physicalism. As we will see later on there are arguments that say that Mary does indeed learn something new but that this does not prove that physicalism is false because that which she learns falls outside the scope of the claims made by physicalism (starting at 3.3.8)[36]. The intuition that if Mary learns something then physicalism is false is what I will call the anti-physicalism inference (API).

[36]There are also argumentations that argue against the practical possibility of Mary's room, however this thought-experiment does not revolve around practical possibilities but about *logical* outcomes [22, p. 3].

3.3.3 What is it about Mary's room that I want to analyze?

The main theme of my analysis of the discussion revolving around the Chinese room (3.2) was about the conceptual confusion regarding the term 'understanding'. The Chinese room was thus implicitly about the *conflict* between first-person and third-person perspectives and the notion of understanding in particular. Mary's room is explicitly about the *relationship* between these two perspectives; this discussion is explicitly about the truth of these two perspectives and particularly about the notion of knowledge.

Physicalism as the claim that all that can be known is physical knowledge, where physical knowledge is knowledge of things in a decomposed and abstract manner, is a claim that can only be made within the third-person perspective because only within it do the decomposed and abstract things exist as described by physicalism. The claim that Mary gained knowledge by experience on the other hand can only be made from within a first-person perspective because such phenomenological knowledge can only be gained by taking a particular perspective. The theme of Mary's room is thus whether or not a third-person description gives rise to knowledge of a first-person kind; it is about the relationship between these two perspectives. It is in this sense that the discussion is explicitly about the relationship between third-person and first-person perspectives unlike the Chinese room.

But it shares a commonality with the Chinese room in that its outcome is also perspective dependent, the view of a perspective from a perspective; the third-person perspective has an idea about the conception of the world found within the first-person perspective and vice versa. For instance, KI can only be accepted from within a first-person perspective; it is how the world works, including the relationship between first-person perspective and third-person perspective, as seen from within the first-person perspective. I will discuss this further in the next section.

3.3.4 What does Mary's room look like from within a perspective?

In the previous section I described the positions regarding Mary's room as perspective dependent. This section will serve as a starting point of clarifying this position by showing what Mary's room looks like when only viewed from one perspective; these positions are of course fully perspective dependent. What Mary's room looks like when not having such a exclusive position will be discussed in the next section.

Within a third-person perspective there is no such thing as phenomenological

knowledge. *All that exists from a third-person perspective is knowledge that is independent of any perspective in particular (contra phenomenological knowledge).* The redness of red as a property that is dependent on a particular perspective can thus from a third-person perspective only be considered as hallucinatory/fiction. Therefore, physicalism in this sense is not incorrect because it already sees the world as physical and nothing more/else[37].

Thus, from a third-person perspective, the knowledge intuition is invalid because talk about phenomenological knowledge is *meaningless*. It then becomes impossible for Mary to learn anything new[38].

Of course from a first-person perspective such a view would be ridiculous. Although the world is indeed physical in the sense that things that can be grabbed are physical, the world is not experienced as consisting of atoms, neurons, functional mechanisms, etcetera. Therefore, from such a perspective the mind can not be understood in such physical terms. Mary *has to* learn something new when going outside her room because she has not really learned anything about the mind or the world (i.e. a world of subjects and non-decomposable objects) in the first place. Whatever third-person material Mary has read (when inside her room) would be *irrelevant*. It is for this reason that the knowledge intuition only exists by virtue of the first-person perspective and is contained within it.

In conclusion, Mary's room is either accepted or rejected if an individual adheres to only one of these perspectives. An example of such a third-person perspective would be eliminative materialism[39]. However, the discussion about Mary's room beyond the exclusion of one of the perspectives is about the relationship between the two perspectives in the relevant sense that both perspectives are said to somehow say something true about the world. I would now like to discuss these non-exclusive positions.

3.3.5 Double vision

What is a non-exclusive position? It is a position that acknowledges first and foremost the existence of both perspectives and gives a certain validity to both of them. Unlike exclusive positions which accept only the validity of one perspective. This was how I analyzed the Chinese room. Such analysis was appropriate because it could highlight conceptual confusion within that context. In the pre-

[37] Physical here can also mean biological, chemical and/or functional.

[38] I hope that it is clear that this is not an argument in favor of physicalism but a mere description. If it would be an *argument* it would be circular; physicalism is true because there is only physical knowledge.

[39] According to Jackson a lot of positions with respect to Marys room are covert forms of eliminative materialism.

vious section I showed what Mary's room would look like with such an analysis. The conceptual confusion that can be found in Mary's room requires a slightly different kind of analysis.

What I would like to propose is that these non-exclusive positions might seem independent from both third-person and first-person perspective but that they are not. This becomes clearest when we consider that acceptance of the Mary's room though-experiment (as rejecting physicalism) on the basis of the knowledge intuition is first-person driven. Accepting the knowledge intuition as true means accepting a first-person conception of the world.

Therefore even if one acknowledges that the world is (or can be conceived as a) physical decomposable thing but one still maintains that the knowledge intuition is correct then one is still influenced by the first-person perspective. Such a non-exclusive position is therefore not perspective independent.

Non-exclusive positions regarding Mary's room exist because it is explicitly about the relationship between third-person conceptions and first-person conceptions. This non-exclusivity may also hide conceptual confusion. It seems to me that there is the assumption that if one explicitly discusses two conceptual domains (e.g. the first-person image and the third-person image) that there can be no conceptual discussion between them. I will start discussing this (hidden) conceptual confusion in section section 3.3.7 but I will first show what a rejection of KI looks like from within the third-person/first-person framework because it is also a perspective-driven position.

3.3.6 Rejecting the knowledge intuition

One reason given in favor of the rejection of the knowledge intuition is that we simply do not know what it means to know all that there is to (physically) know. Therefore our intuitions about knowing all the (relevant) physical facts could simply be misguided. This particular argumentation does not go to the core of the issue though. The knowledge intuition is not based on our ignorance regarding (semi-)omniscience, it is based on the first-person perspective on the world itself.

From a third-person perspective, it is acceptable to say that this intuition does indeed come from our ignorance; we do indeed not know what it means to know everything that there is to know. The acceptability of this claim lies in the third-person assumption that if we would know everything there is to know, we would not require a first-person perspective. If we know all there is to know from a third-person perspective we would have completed the ideals of the third-person

perspective. And when this goal becomes fulfilled, then any epistemology that is based on a non-objective ontology would disappear and thus also its faulty intuitions. However, this argumentation is third-person driven; it assumes that its own ontology is correct. The first-person perspective could then simply reject this argumentation by stating that (only) its own ontology is correct.

It should be pointed out that this reply is not based on a conceptual confusion of specific conceptions but rather from a misconception regarding the nature of the first-person and third-person perspectives. This is not the only possible form of reply against Mary, one could also accept KI but still reject API which will be discussed the following sections.

3.3.7 How could one accept KI while rejecting API?

Third-person critiques of Mary's room (in defense of not rejecting physicalism) have also followed a different route (for instance, the ability hypothesis and the acquaintance hypothesis, will be discussed in the next sections), to accept that the knowledge intuition is true while maintaining that physicalism is not refuted because of it (i.e. denying API). Which will be shown to be problematic.

If one accepts that there is such a thing as a first-person perspective while still holding on to the ideal of an epistemology that is based on not being any perspective in particular, then one can conclude that an illusionary first-person epistemology is distorted by virtue of being based on a *particular* perspective, thus phenomenological knowledge is knowledge from a perspective (to be more precise, a first-person perspective). Such an epistemology (or rather such knowledge) has different properties. The odd thing about this particular defense (i.e. that the knowledge argument is true and so is physicalism) is that it uses a first-person intuition, namely the knowledge intuition. Those that have decided to take this approach are essentially saying that a first-person perspective is correct but that a third-person perspective could logically exist next to it as well. The next sections will go in to more depth regarding this approach starting with the ability hypothesis.

3.3.8 What is the ability hypothesis?

I would now like to focus on one particular type of reply against Mary's room, namely the ability hypothesis. The ability hypothesis states that Mary acquires know-how when leaving her room. It is claimed that this know-how is not the same as knowing that. Knowing-that is propositional knowledge and know-how is not. Physicalism makes a claim about knowing that. All there is to know (i.e.

propositional knowledge) is physical. The ability approach takes the position that personal experience produces an ability and that this ability can not be obtained by knowing-that. Therefore, Mary does indeed learn something but it has nothing to say about the validity of physicalism (which makes a claim about propositional knowledge). This particular reply is one that accepts KI but rejects API.

3.3.9 Two conceptions of know-how/ability

In the previous section I have stated that Mary gains (according to the ability hypothesis) know-how. This know-how is a first-person conceptualization of know-how. This is a conceptual confusion regarding the term 'know-how' because that first-person conception is used to defend physicalism[40].

First, there is a distinction between ascribing knowledge from a first-person or third-person perspective. To say that someone possesses this or that knowledge is the same as ascribing that knowledge; this knowledge can be ascribed by someone else or by the individual itself. I make this distinction between self-ascription and non-self-ascription of knowledge to make clear that this is *not* what is meant by the distinction between ascribing knowledge in a first or third-person manner. The relevant (perspectival) ascriptions are based on how we have conceptualized a person, as an object (i.e. third-person) or a subject (i.e. first-person).

Second, there is the relevant distinction between a first-person conception of know-how and a third-person conception of know-how. Knowing-that on the other hand is conceptualized equally in both perspectives. We know that there are certain unconscious processes which are not directly observable; we also know that trees lose their leaves in autumn which is directly observable. That which is known can be different but both take the same higher-level form of knowledge (i.e. propositional knowledge).

A first-person conception of know-how (as non-decomposable) is something that is only available within a first-person perspective. Such a conception of know-how is always dependent on a *subject* that knows how to do something. Such know-how can thus not be present in a third-person account of the world (including qualia) because its ontology only consist of objects; there are no persons to possess this type of know-how. But from such a third-person perspective it is not denied that individuals posses abilities; for instance, a biological organism

[40]The ability hypothesis does not validate physicalism. It only rejects the anti-physicalist inference.

is still capable of walking (i.e. the ability to walk) despite that individual not being conceptualized as a person (i.e. a first-person conception).

Within a third-person perspective of the world all know-how is decomposable into (unobservable) knowing-that. From a third-person perspective all know-how can be (and must be) decomposed into know-that. Physicalism (as a third-person perspective) does not merely make the claim that all know-that is physical, it also claims by being a token third-person perspective that all that there is to know is know-that.

The idea that all know-how is reducible to know-that is called *intellectualism*. This position has recently gathered more ground, in particular due to the writings of Jason Stanley [36]. He goes against the (once) popular (and accepted) position of *anti-intellectualism* which has taken root by the writings of Gilbert Ryle [30]. Ryle holds that to know how to F is a form of non-propositional knowledge. According to Stanley this is false; know-how is a subspecies of know-that. He defends this position not by giving a philosophical account of the difference proposed by Ryle between non-propositional and propositional mental states, but by giving a semantic analysis of the ascriptions themselves. This is different from my account of the epistemological possibility to reduce know-how to know-that; what I am claiming is that we can describe know-how in terms of know-that if certain ontological assumptions are made (and thus also ascribe it in such a way), as is the case with a third-person perspective. What Stanley provides is a particular defense of such a third-person perspective, using a semantic analysis. Conversely, Ryle takes a first-person stance. I am not claiming that they are correct or incorrect, only that their accounts are perspective depended.

In summary, knowledge is ascribed from a perspective and the form that this knowledge takes depends on that perspective. From a third-person perspective we are only able to ascribe know-how as a decomposable form of knowledge (e.g. know-that). From a first-person perspective we can not. The distinction that I have made is thus not the same as the difference between self-ascription and non-self-ascription; self-ascription is related to it however. When we ascribe knowledge to someone in a first-person perspective we ascribe it in such a way as we would also ascribe this knowledge to ourselves (as subjects).

A concrete example: Consider someone (called John) riding a bike, John has the ability to ride a bike; there are two ways in which we can assign this ability to him. An ability consists of know-how. Know-how can be a third-person conception or a first-person conception. This has a s a consequence that 'ability' is a polysemous term. John can be assigned both the first-person ability and

the third-person ability of riding a bike.

John may know that he has to move his legs in a circular motion while keeping his feet on the pedals and that in order to keep his balance he can not move his body too much to the sides of his bike. From his own first-person perspective he does know-that to riding a bike and by looking at him ride a bike we can also assign such first-person conception of riding a bike to him. The first-person conception of the ability to ride a bike consists of nothing more then observable procedures contrary to a third-person conception.

For instance, he does know-how to move his legs and keep his balance but he is unable to tell us how he does this in order for us to be able to do this as well; he is unable to express what we need to do (besides moving our legs in a certain way, keeping balanced, etcetera.) on a lower level decomposition (e.g. physical or biological). From a third-person perspective we would be able to tell what these observable procedures are composed of .We could therefore also be able to assign the third-person conception of ability to him despite it being a distinct conception.

The ability hypothesis is based on the distinction between know-how and know-that as being different types of knowledge. This type of reply can only be made from a first-person perspective because only within that perspective is there such a difference between know-how and know-that. But this conception of abilities is then used to defend a position in which such distinction does not exist. It is the application of a conception of abilities in a domain where it does not belong. Therefore, the ability response is based on a conceptual confusion regarding abilities.

3.3.10 Why would experience produce abilities?

In the previous section I conclude that the ability response is based on a conceptual confusion regarding abilities. I would now like to strengthen this conclusion by showing why one would think that the experience of red produces an ability. I will conclude that such an idea is based on a first-person conception of experience and its correlations made with a first-person conception of abilities.

When we think of ourselves as individuals with certain abilities we do not often think about the underlying mechanisms of these abilities. Within our first-person experience of the world all that there is to these abilities are the observable procedural mechanisms. This description of abilities is a first-person conceptualization of abilities which stands apart from a decomposed conceptualization of these abilities (i.e. a third-person conceptualization). The same

applies to the learning process of such an ability; from a first-person perspective we learn by non-decomposable experience.

Because of this relationship between experience and ability we may also think that the experience of red itself produces an ability. This is why the ability hypothesis claims that the experience of redness enabled Mary to imagine (which is an ability) other people experiencing redness. This also means that the ability hypothesis is grounded within the first-person perspective; it is based on both a first-person conception of abilities (as shown in section 3.3.9) and first-person correlative relationship between these abilities and a first-person conception of experience.

3.3.11 What is the acquaintance hypothesis?

Another reply which sees the knowledge intuition as correct but does not adhere to the anti-physicalist inference, is the acquaintance hypothesis. According to this approach, when Mary is released from her room she becomes acquainted with the (phenomenological) property red. This acquaintance results in non-propositional knowledge just like with the ability hypothesis. What exactly acquaintance is is notoriously vague but I will use Conee's description that knowing something by acquaintance "requires the person to be familiar with the known entity in the most direct way that it is possible for a person to be aware of that thing" [5, p. 144]. This is another way of saying that knowledge by acquaintance occurs when directly observing an object, which is the only way from which one can observe things within a first-person perspective. The description given by Conee is thus strongly related to my description of the first-person perspective.

Why would such direct observation lead to non-propositional knowledge? Because what Mary is ignorant about is not a fact but a property, namely the property red. This is expressed by Conee as "To come to know a property is to become acquainted with the property, just as to come to know a city is to become acquainted with the city, and to come to know a problem is to become acquainted with the problem." [5, p. 140]. From a first-person perspective such relationships of acquaintance are acceptable. For instance, from a first-person perspective it would be acceptable to say that John knows Bill because John has met Bill in some direct way and not merely read some description of her, which is another way of saying that John has first-hand knowledge of Bill.

In the case of Mary, she has not had any first-hand knowledge of the quale red (which is a property) but she does know everything descriptive about red. It

is than claimed by the acquaintance hypothesis that such first-hand knowledge can not be inferred from descriptive/propositional knowledge (cf. John has read everything about Bill but that does not mean that he knows Bill (by acquaintance)).

One could then ask, like Gertler, "Why is it that, in contrast to other objects of knowledge, we can know (or effect an 'optimal cognitive accomplishment' with respect to) phenomenal qualities only by experiencing them?"[15, p. 327]. She also notes that such distinction between phenomenal qualities and physical qualities underlies an ontological distinction. I would like to point out that this ontological distinction does not exist within a third-person perspective. It is for this reason that a first-person conception of acquaintance is used in order to defend a third-person position.

The acquaintance hypothesis is in this respect similar to the ability hypothesis. However there does not appear to be a third-person conception of acquaintance with which this first-person conception can be confused. In other words 'acquaintance' does not appear to be a polysemous term.

But here too one should see that if a third-person perspective wants to be both complete and abstracted away from any first-person perspective; it would also require some account of acquaintance itself. Then acquaintance becomes a yet-to-be decomposed functional process, which would be expressed in terms of know-that.

3.3.12 Where do the acquaintance hypothesis and ability hypothesis go wrong?

Both of these approaches do something similar; both affirm that the knowledge intuition is correct. This knowledge intuition is a first-person intuition. It follows from a first-person perspective. Both then use first-person concepts of knowledge; know-how and knowledge by acquaintance, to show that physicalism is not refuted by it. Both use these first-person conceptions in order to defend a third-person account of the world (e.g. physicalism).

You can not save a position (e.g. physicalism) with something (e.g. acquaintance or know-how) if that something is conceptualized differently (e.g. objective acquaintance and objective know-how) within that position. Only from a first-person perspective do acquaintance and know-how stand apart from know-that but from a third-person perspective they do not and therefore they can not be used to save a third-person perspective.

3.3.13 And lastly: What about Jackson's epiphenomenal qualia?

A first-person conceptualization of qualia does not fit with a third-person conceptualization of the world; the redness of red is not a part of it. Yet it does contain knowledge about the world, just a different (ontological) conception of the world. So one could then claim that physicalism is incorrect if one already accepts that there is also (or only) a world of subjects (and non-decomposable objects). Jackson then makes the odd jump to say that qualia must then be epiphenomenal, but this too misses the mark; a first-person conception of qualia is indeed epiphenomenal when placed in a third-person conception of the world. A first-person conception of qualia is non-decomposable and thus can not act in a world that is decomposed. But this is a form of ontological misplacement. Granted, once placed in the setting of a decomposable world, it would not act; that does not mean that it could be placed there *in the first place*.

3.3.14 Mary's room as a case of conceptual confusion

I have described Mary's room as a thought-experiment that is about the relationship between the third-person and the first-person perspective but also in which the positions regarding that relationship is perspective dependent. From a first-person perspective third-person knowledge will not produce the information contained in qualia; hence the knowledge intuition. The third-person perspective on the other hand does not grant qualia any special properties and therefore whatever information that is contained within qualia is reducible to some third-person description which takes the form of propositional knowledge.

Even when people do not adhere to either a third-person perspective or a first-person perspective their position is still perspective dependent. This can be seen when analyzing the ability hypothesis. This response stems from a conceptual confusion regarding ability/know-how. It uses a first-person conceptualization of ability to defend a third-person position in which such a first-person conception of ability is seen as false.

Discussion/conclusion

Artificial intelligence as a third-person scientific field.

Artificial Intelligence is a third-person science, there are different paradigms within Artificial Intelligence but all require a particular decomposition of the mind. This decomposition of the mind is a decomposition of the mind as seen from a first-person perspective. As a consequence, we find the same problem that we have encountered during the analysis of philosophical thought-experiments (chapter 3), namely a conceptual confusion between first-person conceptions and third-person conceptions. This problem is more severe when we talk about understanding the mind/intelligence, which is the cognitive goal of the project of Artificial Intelligence. Then questions arise regarding the nature of the mind and how we should understand it. Conversely, a more engineer-wise take on Artificial Intelligence does not have such big problems. Whatever one's goals, what should always be kept in mind is that the object of study (or replication) is third-person conception and not a first-person conception.

However, questions are asked about whether a machine can actually feel, can actually understand, can actually be intelligent etcetera. I purposely used the word 'actually' because it shows what is misleading with respect to these questions, it presupposes that what it means to (actually) feel has to be conceptualized as a first-person conceptualization. This is only true from a first-person perspective itself. Since Artificial Intelligence is a third-person science it does not require to take such a first-person stance. As a consequence, such (first-person) questions are irrelevant to its undertaking.

Furthermore, such questions (and similar ones) arise from a conceptual confusion and in effect diminish the independence of Artificial Intelligence as a scientific field required for any proper scientific field; AI becomes reliant on ideas that it has no control over. Such conceptual confusion about the mind stems from the use of metaphors (2) which create two distinct perspectives (1) which we are not always aware of (most importantly due to the existence of polysemous terms (2.3.2)). However, AI as a strictly third-person perspective, can answer these questions third-person wise; it can provide a functionalistic description of feeling/understanding/intelligence etcetera.

I should add that not everyone categorizes AI as *only* a scientific field, some see AI as something more broadly defined including for instance the philosophy of AI. Arguments given for this view are: 1) If you would present a list of fields, for instance law, biology, sociology, and AI, people would in general say that AI is best equipped to answer these first-person questions. 2) Students and scientists

of AI often ask and debate these first-person questions. 3) Third-person research is in part fueled by these questions. One could counter-argue against arguments 1) and 2) that this way of looking at AI is simply a result of the conceptual confusion I have addressed. Argument 3) is more of a practical point; it is indeed paradoxically true that conceptual confusion has led to more interest and (scientific) research into AI; but it has no implications beside that[41]. However, one could still maintain that AI includes more than just AI as a science, in that case the points made with respect to AI as belonging to the third-person perspective hold to the extent that one considers AI to be only a scientific research project (or a particular sub-field of AI).

Strong AI and weak AI

I would now like to say something about the distinction between strong AI (i.e. AI that has all the capabilities that humans have) and weak AI (i.e. AI with lower capabilities compared to humans), I will argue that this distinction is misguided.

If the distinction between weak and strong AI is about the discussion about whether or not the mind is a machine is merely metaphorical, then the discussion is simply misguided. If the discussion is about whether the metaphor used (MIND IS A MACHINE or MIND IS X) is appropriate one should distinguish between appropriateness within a third-person perspective and between the third-person and first-person perspective. Within a third-person perspective we may not have found an appropriate metaphor yet; we have not found a metaphor that qualifies as both accurately descriptive and intuitively appealing; this is simply a part of the exploratory nature of new scientific fields and research projects. This internal (third-person) conflict should be distinguished from a clash between third-person and first-person perspectives regarding the mind. Because of the nature of these two perspectives any objective account of the mind will be considered as using inappropriate metaphors from a first-person perspective.

If the distinction between strong and weak AI is about whether or not an artificial entity can 'actually' feel, understand, have mental states with intentionality etcetera then it becomes a matter of what perspective is taken. Those who take the position that weak AI is correct are usually looking at AI from a first-person perspective but still demand a third-person account of the mind/AI; this can not

[41] To be clear, I am not saying that proper first-person philosophical investigation has no relevant consequences on third-person science. In order to properly model a first-person phenomenon it is important to first grasp (first-person wise) what that (manifest) phenomenon is.

be done, not because AI can not have all these qualities described from a third-person perspective but because any third-person account will not correspond with a first-person perspective[42].

Conclusion

In chapter 1 I described the two perspectives which I claimed consisted of distinct conceptions (in particular about the mind) which were sometimes used as if belonging to the other domain; third-person conceptions about the mind were used as if they were first-person conceptions and vice versa. To show how this conceptual confusion could arise I described the metaphorical relationship between these two perspective in chapter 2. I concluded that third-person conceptions were metaphorical conceptions using first-person conceptions. To be more precise a third-person metaphorical conception has the structural form A FIRST-PERSON CONCEPTION IS X. This could at times be problematic because both third-person conception and its related first-person conception are referred to with the same term. For instance, 'mind' can refer to both a first-person conception and a third-person conception of the mind. I showed that this was indeed the case by analyzing two thought-experiments with this framework in chapter 3.

The first thought-experiment I examined was the Chinese room. In it I found a conceptual confusion with respect to the conception of understanding. Searle adhered to a first-person conception of understanding while his opponents adhered to a third-person conception of understanding. Both criticized each other using these distinct conceptions. Searle set up the Chinese room as analogous to a computer. THE MIND IS A COMPUTER is one possible third-person conception of the mind. Searle claimed that the Chinese room showed that the mind can not be a computer because there was no understanding of Chinese. This is correct from a first-person perspective because within such a perspective understanding can not be decomposed or abstracted. The capability of the mind to understand is conceptualized in a first-person manner when the mind itself is a first-person conception, that is to say when it becomes an abstract and decomposed conception. The Chinese room is an analogy of such a third-person conception. Therefore Searle will always be right when he says that symbol manipulators are unable to understand Chinese (or anything else), but only because he take a first-person perspective.

His opponents on the other hand take a third-person perspective. From such a perspective it possibly to say that understanding (as a third-person conception)

[42] As far as I can see AI as a scientific field will never be able to provide such an account.

does indeed occur. Some might argue that it is the system as a whole that does the understanding, which they declare as a refutation of Searle's position (that the Chinese room can not understand). But this misses the point that Searle takes a first-person perspective and therefore no modification of the Chinese room (resulting in yet another decomposed and abstract entity) would refute Searle's position. The intend of these replies to *refute* Searle position follow from the conceptual confusion between first-person image and third-person image. It is therefore no surprise that Searle dismisses this reply (and similar replies) by showing that such a modification can not understand. But he too participates in conceptual confusion by dismissing such replies as if they were concerned with a first-person conception of understanding. In short, *the discourse surrounding the Chinese room (as far as I have analyzed it) is based on a conceptual confusion between first-person conceptions and third-person conceptions.*

The other thought-experiment that I analyzed was Mary's room. This particular thought-experiment proved to be more difficult to analyze because it is explicitly about the third-person and first-person perspectives. It dealt with the relationship between the two. Jackson intended Mary's room to be an argument against the idea that all information is physical information. He did so by contrasting physical knowledge with phenomenological knowledge. Jackson claimed that Mary's room showed that such phenomenological knowledge can not be reduced to physical knowledge because Mary learned something new. In the light of my third-person/first-person framework such physical knowledge is produced from a third-person perspective whereas phenomenological knowledge required a first-person perspective. It is in this sense that Mary's room is explicitly about the relationship between first-person and third-person perspective. *One would thus not expect to find any conceptual confusion between these two perspectives within this context.*

Nevertheless, I did find conceptual confusion within the ability hypothesis. The ability hypothesis argued that Mary did learn something new but that this was no problem for physicalism because Mary acquired know-how and physicalism only made claims about factual knowledge. I showed that there are first-person conceptions and third-person conceptions of abilities. A first-person conception of abilities only includes observable procedures whereas a third-person conception included possible unobservable mechanisms. It is for this reason that from a first-person perspective that there are things that one simply knows (in terms of irreducible know-how). Hence the distinction between know-how and know-that. From a third-person perspective abilities are not simply things one knows how to do. Such abilities are physically grounded and can thus be expressed in purely factual terms. It is in that sense that from a third-person perspective

know-how can be reduced to know-that. This is were the conceptual confusion becomes apparent. The ability hypothesis defends a third-person conception of the world, physicalism, by make use of a first-person conception of the world. If all know-how is reducible to know-that, as is the case from a third-person perspective such as physicalism, then it becomes impossible to claim that Mary acquires know-how which is irreducible to know-that.

In conclusion, both the Chinese room and Mary's room showed cases of conceptual confusion. Such an analysis using the third-person/first-person framework could also by applied to other thought-experiments, like Davidson's swampman and Putnam's twin earth. It has been argued by some that because swampman lacks a casual history we can not attribute any thoughts to him. One could ask if such a claim can be made from within a first-person perspective, considering that from a within a first-person perspective such a casual history does not exist. One would expect the outcome of Putnam's twin earth to be perspective-dependent considering that H_2O and XYZ do not exist within a first-person framework.

I was motivated to write this thesis by the difficulty that I had communicating with people regarding certain concepts (in particular first-person concepts). One could now ask if this has been improved. By and large the answer is yes. Although, to be honest, at first some problems arose with explaining the third-person/first-person framework itself; it was not all that clear how I viewed this distinction. When the distinction was made clear it was deemed somewhat trivial. However, after discussing certain thought-experiments (the Chinese room in particular) the third-person/first-person framework still proved helpful even though I and others became aware of the crudeness of this framework.

This framework proved helpful with respect to the Chinese room, where it was found that the systems reply was indeed a third-person response, but that it was a response to the notion that the man in the room could understand Chinese from a third-person perspective. If you take the book away he would not be able to understand Chinese. As a result whatever there is that understands Chinese cannot just be the man himself, not from a third-person perspective and therefore also not from a first-person perspective either. This of course does not imply that the system as whole understands Chinese from a first-person perspective but it does reject the Chinese room as an argument against strong AI because it leaves open the possibility that the system as a whole understands Chinese from a third-person perspective. In short, the third-person/first-person framework and its terminology proved helpful with this discussion.

Some problems still arose however. Some people apparently experience the

world very differently from how I do and how I thought others would do as well. For instance I met one person who claimed to (literally) experience the world as a dynamic system, which seems a bit too odd to me. Another person claimed that her experience of black is that not of a colour since black does not reflect light and therefore is not a colour. Even after explaining that this is a third-person description of black not being a colour, she still maintained that she did not experience black as a colour (and for that third-person reason as well). I am going to assume that this were indeed their first-person experiences, and since I can not dictate what people experience, my current third-person/first-person framework must have been (in those cases) somewhat too crude in order to provide the communicational clarity that I was aiming for.

It may be that people explain their experiences in third-person ways even though it is not exactly what they are experiencing or that their third-person knowledge has a great effect on their actual first-person experience. These follow-up questions are related to one of the things that puzzled me during the writing of this paper; how has our current first-person perspective (analogous to Sellars' current manifest image) been shaped; to what extent is it malleable and shaped by a third-person perspective? And if the effects of the third-person perspective are strong then what does this say about the third-person/first-person framework as I have currently sketched it.

References

[1] Bailer-Jones D.M., (2002). Models, metaphors and analogies. In: P. Machamer, M. Silberstein, eds. *The Blackwell guide to the philosophy of science.* Malden, MA: Blackwell Publishers. Ch.6.

[2] Block, N., (1980). Introduction: What Is Functionalism? In: *Readings in Philosophy of Psychology.* Cambridge, MA: Harvard University Press.

[3] Brook, A., (1997). Reconciling the two images, In: S. Nuallain, P. Mc Kevitt, E. Mac Aogain, eds. *Two sciences of mind.* Philadelphia, PA: John Benjamins North America. p. 299-310.

[4] Brown, R.H., (1977). *A Poetic for Sociology: Toward a logic of discovery for the human sciences.* The University of Chicago Press.

[5] Conee, E., (1994). *Phenomenal Knowledge,* Australasian Journal of Philosophy 72: p. 136-150.

[6] Davidson, D., (1987). *Knowing One's Own Mind,* Proceedings and Addresses of the American Philosophical Association, 60 (1987), 441-58.

[7] Dennett, D., (1984). *The Role of the Computer Metaphor in Understanding the Mind,* Annals of the New York Academy of Sciences Volume 426, Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer p. 266-275, November 1984.

[8] Dennett, D., (1984). *Elbow Room*, New York: Oxford University Press.

[9] Dennett, D.C., (1995). Intuition pumps. In: J. Brockman. *The third culture.* New York: Simon and Schuster.

[10] Dennett, D., (1996). *Cow-sharks, Magnets and Swampman*, Mind and Language, 11(1), p. 76-77.

[11] Dilthey, W.m (1914-) Gesammelte Schriften, 18 vols. Stuttgart: B. G. Teubner (vols 1-12); Gttingen: Vandenhoek and Ruprecht (vols 13-18)

[12] Dooremalen, H., de Regt, H., and Schouten, M., (2007). *Exploring Humans: An Introduction to the Philosophy of the Social Sciences.* Amsterdam: Uitgeverij Boom.

[13] Eliasmith, C. (2003). *Moving beyond Metaphors: Understanding the Mind for What it Is*, Journal of Philosophy, 100(1), p. 493-520.

[14] Frigg, Roman; Hartmann, Stephan, "Models in Science", *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/spr2012/entries/models-science/.

[15] Gertler, B., (1999). *A Defense of the Knowledge Argument*, Philosophical Studies, 93, p. 317-336.

[16] Giere, R,N., (2006). *Scientific Perspectivism.* Chicago, IL: University of Chicago Press.

[17] Graham, G., "Behaviorism", *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2010/entries/behaviorism/.

[18] Jackson, F., (1982). Epiphenomenal Qualia, *Philosophical Quarterly*, 32, p. 127-36.

[19] Jackson, F., (1986). What Mary Didn't Know, *Journal of Philosophy*, 83, p. 291-295

[20] Jaynes, J., 1978. *The origin of consciousness in the breakdown of the bicameral mind.* Boston: Houghton Mifflin.

[21] Lakoff, G., and Johnson, M., (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

[22] Ludlow, P., Stoljar, D., and Nagasawa, Y., (2004). *There's something about Mary,* Cambridge, MA: MIT Press.

[23] McAllister, J.W., (2002). Historical and Structural Approaches in the Natural and Human sciences. In: J.W. McAllister, J. van Benthem, A. Rip, H. Philipse, eds. *The Future of the Sciences and Humanities,* Amsterdam: Amsterdam University Press, Ch.2.

[24] Morrison, M. and Morgan, M.S., (1999). *Models as Mediators: perspectives on natural and social science,* Cambridge: Cambridge University Press.

[25] Morrison, M. and Morgan, M.S., (1999). Models as mediating instruments, In: M. Morrison, and M.S. Morgan, eds. *Models as Mediators: perspectives on natural and social science* Cambridge: Cambridge University Press. Ch.2.

[26] Morrison, M., (1999). Models as autonomous agents, In: M. Morrison, and M.S. Morgan, eds. *Models as Mediators: perspectives on natural and social science* Cambridge: Cambridge University Press. Ch.3.

[27] Nagel, T., (1979). *Mortal Questions.* Cambridge: Cambridge University Press.

[28] Nagel, T., (1986). *The View From Nowhere.* Oxford University Press.

[29] Northoff, G., and Musholt, K., (2006). How can Searle avoid property dualism? Epistemic-Ontological Inference and Autoepistemic Limitation, *Philosophical Psychology,* 19(5), p. 1-17.

[30] Ryle, G., (1949). *The concept of mind*, New York: Barnes and Noble, Inc.

[31] Schoemaker, S., (1994). The First-Person Perspective. *Proceedings and Addresses of the American Association,* 68, p. 7-22.

[32] Searle, J., (1980). Mind, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), p. 417-457.

[33] Searle, J., (1992). *The Rediscovery of the Mind,* Cambridge, MA: MIT Press.

[34] Searle, J.R., (2002). Why I am not a property dualist. *Journal of Consciousness Studies,* 9(12), p. 57-64.

[35] Sellars, W., (1962). Philosophy and the Scientific Image of Man. In Frontiers of Science and Philosophy, edited by R. G. Colodny, p. 5-78. University of Pittsburgh Press, Pittsburgh.

[36] Stanley, J. and Williamson, T., (2001). "Knowing How", *Journal of philosophy,* 98, p. 411-444.

[37] Suérez, M., (1999). The role of models in the application of scientific theories, In: M. Morrison, and M.S. Morgan, eds. *Models as Mediators: perspectives on natural and social science.* Cambridge: Cambridge University Press. Ch.7.

[38] Taylor, J.R., (1995), *Linguistic Categorization: Prototypes in Linguistic Theory,* Oxford: Clarendon Press.

[39] Van Gulick, R., (2004). So many ways of saying no to Mary. In: P. Ludlow, D. Stoljar, and Y. Nagasawa, (2004). *There's something about Mary,* Cambridge, MA: MIT Press

[40] Watt, S. N. K., (1997). The Lion, the Bat, and the Wardrobe: Myths and Metaphors in Cognitive Science. In: S. Nuallain, P. Mc Kevitt, E. Mac Aogain, eds. *Two sciences of mind.* Philadelphia, PA: John Benjamins North America. p. 299-310

[41] Wilkes, K.V., (1988). *Real People: Personal Identity without thought-experiments.* Oxford: Clarendon Press.

[42] Wilson, T., Forthcoming, Dewey, Nagel and the Problem of Objectivity.