DOING THINGS IN YOUR HEAD:
A PHILOSOPHICAL ESSAY ON MENTAL ACTION

by

Niko Eugenia Scharer

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright by Niko Eugenia Scharer (1997)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Abstract:

John B. Watson distinguished between the objective study of human behavior and the subjective study of consciousness. Like Watson I hold that human behavior is the subject of psychology; but there are problems with Watson's distinction. First, most observable human behavior is itself conscious; and second, some conscious behavior does not involve overt movement at all. Many behaviors, indeed many actions, are performed 'in the head'.

I advance a theory in which behavior is a single general kind; mental and overt behavior are merely two interestingly distinct forms. Most behavior depends on prior events in which the behaver learns from sensory experience. Thus I develop a naturalizing theory of representation (modeled on artistic representation) which treats the brain as a medium in which representational structures are formed when we interact with the world around us. These representational neuro-structures are utilized in the production of behavior. The resulting mental and overt behaviors are likewise representational and can be interpreted as meaningful.

The medium theory of representation provides a framework in which to understand mental and overt behaviors. Mental behaviors differ from overt behaviors in that they involve the active suppression of bodily movement. Suppression is, in my theory, central to the distinction between behavior and action; and thus, to my account of rudimentary action and the closely related mental behavior of desiring, and to more complex forms of action, including 'situated' behaviors, speech acts, deliberate action and mental acts.
Finally I turn to the things we do 'in our heads', and the varied consequences of the different ways and circumstances in which we perform these mental acts. The consequences that are of greatest interest concern our conceptions of agency and the mind. Here it is of central importance that we learn to internalize behaviors within the context of the overt practices and activities that constitute a society. This approach to human behavior yields a picture, not of a mind, but of a mental life which can take on many forms. The dominant forms of our mental life are best understood within the context of the social environment in which we develop.
ACKNOWLEDGMENTS

There are many people who made my doctoral studies at University of Toronto enjoyable and rewarding. I take this opportunity to thank my professors, friends, fellow students and the administrative staff at the Philosophy Department who supported me in so many various ways — through lectures, discussions and arguments, evenings of great food and company, advice about everything from revisions to changing printer cartridges, and most importantly with friendship and emotional support.

I am most indebted to my supervisor, Ian Hacking. I could not ask for a more stimulating, generous and supportive supervisor. My dissertation benefited greatly from Ian's comments and criticisms, and also, less directly, from his own philosophical work and teaching.

I am also indebted to my advisor, William Seager. He provided excellent criticisms and challenging discussions.

I also want to thank Ronald De Sousa, Danny Goldstick, Sonia Sedivy and Stephen Yablo, who read parts of this thesis at various stages of development and discussed central ideas with me.

I am grateful to many at the Philosophy Department of the University of North Carolina at Chapel Hill, but especially to William Lycan. I greatly benefited from my year in Chapel Hill, 1995-1996, where I was always made to feel welcome.

In the first year of my doctoral studies, I was supported by a Connaught Scholarship from the University of Toronto, 1991-1992. I was further supported in my research by the Social Sciences and Humanities Research Council of Canada, 1992-1995.

Finally I would like to thank my family, without whose love, support and encouragement I could not have written this thesis. Thanks go to my mother, Iris, my sisters, Ilona and Margo, and my father, Jeno. I also thank my step-children, Meg and Campbell, two great kids who were patient and encouraging throughout.

Most of all, I thank my husband, Brad, to whom I dedicate this work.
## CONTENTS

Chapter:

1. CONSCIOUS BEHAVIOR AND MENTAL ACTION: AN INTRODUCTION . . . 1

   A *Prima Facie* Case for Mental Actions and Behaviors (4)
   One Theory of Human Behavior (17)

### PART I:
THE MEDIUM THEORY OF REPRESENTATION

2. PEOPLE AND OTHER MEDIA . . . . . . . . . . . . . . . . . . . . . . . 24

   "I am a camera." (44)

3. USING THE MEDIUM THEORY TO ACCOUNT FOR NEURAL
   REPRESENTATIONS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

   Single-Use, Single-Exposure Media (54)
   Misrepresentation (64)
   Determining the Represented Object (73)

4. MEANING AND CONTENT: FROM PUBLIC LANGUAGE TO PERSONAL
   MEANING . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79

   Meaning and Content (82)
   Personal Meaning (103)
   Twin Earth Considerations (121)

### PART II:
MENTAL AND OVERT BEHAVIOR AND ACTION

5. MENTAL BEHAVIOR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 138

   Conscious Experience and Behavior (140)
   Mental Behavior and the Suppression of Overt Movement (148)
   Internalized Speech (153)
   Desiring (156)

6. DON'T JUST DO SOMETHING, STAND THERE!
   (SUPPRESSION AND ACTION) . . . . . . . . . . . . . . . . . . . . . 168

   Does it Act? (168)
   Habit Formation (174)
   Speech Acts (177)
   Deliberate Action (185)

7. DOING THINGS IN OUR HEADS . . . . . . . . . . . . . . . . . . . . . . . 193

8. MENTAL LIFE . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 210

REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 227
Chapter 1

CONSCIOUS BEHAVIOR AND MENTAL ACTIONS

Two opposed points of view are still dominant in American psychological thinking — introspective or subjective psychology, and behaviorism or objective psychology. Until the advent of behaviorism in 1912, introspective psychology completely dominated American university psychological life.

The conspicuous leaders of introspective psychology in the first decade of the twentieth century were E.B. Titchener of Cornell and William James of Harvard. The death of James in 1910 and the death of Tichener in 1927 left introspective psychology without emotional leadership. Although Tichener’s psychology differed in many points from that of William James, their fundamental assumptions were the same. In the first place, both were of German origin. In the second place, and of more importance, both claimed that consciousness is the subject matter of psychology.

Behaviorism, on the contrary, holds that the subject matter of human psychology is the behavior of the human being. Behaviorism claims that consciousness is neither a definite nor a usable concept. (Watson 1930:1-2)

In the opening paragraphs of Behaviorism, John B. Watson makes a distinction between the subjective study of consciousness and the objective study of human behavior— a distinction which remains central in psychology and in philosophy of mind. Watson’s concern is for developing a more exact, experimental method in empirical psychology. Behavior, but not consciousness, can be observed, measured and studied by other people, including the experimental psychologist. Yet despite the obvious methodological advantages, there is something artificial in this distinction. Most human behavior is conscious behavior. I do not mean that we are conscious of our behavior (although we often are) but that behavior involves consciousness. The vast majority of interesting human behaviors seem to require consciousness for their performance. Even in those rare instances (sleepwalking, automatism, etc.) during which a person behaves in quite complex

---

ways when not fully conscious, the person's behavior is dependent on having previously been conscious — if only when learning how to walk, talk, get dressed, recognize people, fire a gun, etc. A study of human behavior is a study of conscious behavior — of the kinds of things we do when we are, or have been, conscious.

In this essay, I explore the implications of rejecting Watson's distinction, while accepting his starting point: the subject matter of human psychology is the behavior of the human being. But, unlike Watson, I take consciousness to be a central feature of human behavior — we don't just observe people's behavior, we observe them behaving consciously. It is of even greater interest, and in starker contrast to Watson, to reflect on a further fact: not only are most observable, overt human behaviors conscious behaviors, but many conscious behaviors do not involve overt movement at all. Many behaviors, indeed many actions, are performed 'in the head'.

Although much contemporary philosophical discussion has been focused on the relation between the mind and our bodily or overt behaviors, there is little focus on mental behavior. When mental behaviors have been talked about, for instance, in the works of Aristotle, Augustine, Kant, Locke, Descartes, and more recently in Searle's Rediscovery of the Mind (1992), they are explained within the context of some particular account of the mind or of the soul. These accounts typically begin with the premise that the mind or the soul has such a nature or form as the philosopher is advocating, either in itself or in its relation to the brain or body; it then proves possible to account for mental acts such as 'deliberating', 'willing', 'judging', 'intending' and 'imagining'. Thus, discussions of mental behaviors are often inextricably entangled with contestable ontological commitments with respect to mental phenomena or even mind-stuff.
This type of approach has been, especially in the twentieth century, quite rightly criticized. There are countless arguments directed against the 'Cartesian myth', 'Folk Psychology', the existence of mind-stuff, etc.\(^2\) In general, these attacks are quite well grounded — none of these models of the mind or soul has proven satisfactory, whether they come from Aristotle, Plato, Descartes, Kant, Locke, William James, or others. This has led many philosophers to reject altogether, as Watson does, the relevance of mental phenomena to explanations of human behavior. At one extreme, philosophers eliminate mental phenomena from explanations of human behavior.\(^3\) Some are willing to allow a role in explanation for terms that typically refer to mental phenomena, but only provided these terms are reduced to, identified with or explained in terms of neural states, functions or processes.\(^4\) Others prefer to separate the two kinds of explanation, taking an instrumentalist or pragmatic approach to statements about mental phenomena — they are useful ways of predicting and describing human behavior, but they need not entail any ontological commitments to mind stuff at all.\(^5\) These approaches, too, have left many unsatisfied.

In contrast, I start from the assumption that a study of behaviors that are performed 'in the head' or 'with the mind' can provide new insight into the nature (or natures) of the mind or of mental life. One approach to a study of the human body is to investigate how


\(^3\) For instance, Paul Feyerabend (1963), Richard Rorty (1970), Patricia Churchland (1986), Paul Churchland (1981, 1988), and also Stephen Stich (1983) presents arguments in favor of eliminativism without committing himself to it.

\(^4\) For instance, see identity theorists such as U. T. Place (1956), J. J. C. Smart (1962), and Marvin Minsky (1988). Functionalists also hold a physicalist position; however, as mental functions are defined in terms of abstract causal (or functional) roles within the information processing system, reduction (which would demand a one-to-one correspondence between mental functions and brain processes) is not necessarily going to result. See, for instance, William Lycan (1981, 1987), David Lewis (1983).

\(^5\) Daniel Dennett (1987).
we perform bodily behaviors (walking, focusing our eyes, talking, throwing balls, etc.) and in what circumstances. I want to extend this approach to mental behavior, investigating the kinds of things we do 'in our heads', how we do them and in what circumstances. This approach yields a slightly different picture of the mind, or as I will characterize it, of mental life.

To this end, I advance a theory in which behavior is a single general kind, of which the mental and the overt are merely two interestingly distinct forms. But before embarking on this, I present a *prima facie* case that those things that we do in our heads are behaviors and actions, on par with overt behaviors and actions. I am not claiming that we *ought* to treat mental behaviors as we do overt behaviors. Rather, I point out that we already do treat them in much the same way.

**A Prima Facie Case for Mental Actions and Behaviors**

'Action' usually refers to an event involving bodily movements (including those involved in overt speech) which we attribute to an agent: signing a cheque, firing a gun, making a promise, running a marathon, throwing a ball, telling a story, washing the dishes, turning off the light. Yet, we also attribute to agents events that do not necessarily involve any overt bodily movement: calculating an answer, plotting revenge, deciding what to wear, praying silently or imagining what a room will look like painted blue. These are also things that people do. But people do these things 'in their heads' or 'in their minds' rather than through overt bodily movement. Without, at this point, considering to what extent these acts are activities of the brain and thus, in that sense bodily, they can be clearly distinguished from ordinary overt, bodily actions. These are what I call mental actions and behaviors.
What exactly it is for an event to be an action is, of course, a central and difficult philosophical question; my detailed discussion of it is deferred to chapters 5 and 6. What I will show here is that certain kinds of mental events, like those listed above, have more in common with overt actions than with agentless events, and can properly be called actions. One important similarity (though there are others) is the isomorphism between the taxonomies we employ for overt and mental actions. The only significant difference between mental actions and overt actions is that overt actions involve overt bodily movements and mental actions do not; and hence mental acts, unlike overt acts, cannot be directly observed by others.

I am not claiming that every mental event is an action; but neither is every bodily event an action. A vast number of things that take place in our minds are mere events, just as many bodily occurrences are. Muscle twitches, tics, flinching, sleepwalking, having your body manipulated by an external agent and epileptic fits, are not generally things we do with our bodies, but things that happen to us. They are often out of our control, involuntary or automatic. Similarly, mental phenomena such as hearing, hallucinating, having a thought occur to you, or dreaming are not, generally, things that we do voluntarily, but things that happen to conscious people.

In our everyday speech we distinguish between things we do and things that happen to us; and also, between mental events of which we are the agents and mental events that happen to us. Consider the difference between, on the one hand, “Tanya threw the ball” and “Tanya calculated the answer”, and on the other hand, “the ball hit Tanya” and “a thought suddenly struck Tanya.” In the first two sentences, under normal interpretation, Tanya did something, but in the latter two, we describe something that happened to her. Why do the first two sentences, and not the second, appear to describe
actions? The simplest answer is the grammatical one: Tanya is the subject of a verb in the first two, but the object in the latter two. As the subject of the verb, she is doing something; but as the object, whatever is happening is happening to her. But, this easy grammatical answer is insufficient. In the sentence, "Tanya fell", Tanya is the subject, yet she is not, under normal conditions, acting, although she is doing something, i.e. falling. What matters is that the verb describes an action of Tanya’s.

Consider our assessments of voluntariness. Aristotle’s classic treatment of the problem still captures central features of our concept. In the first sentence of Book III of the *Nicomachean Ethics* he states that voluntary feelings and actions are subject to praise or blame, and those that are involuntary are pardoned or even pitied. Whether or not an agent is subject to praise or blame for an action is one indication that we are treating the event as the agent’s voluntary action. We hold people responsible when we judge that they act voluntarily. We do, of course, also praise or blame people for not acting in certain circumstances; but I will leave acts of omission and the closely related acts of ‘standing fast’ for later discussion, dealing first with more straightforward voluntary action.

With respect to praise and blame, calculating an answer is like throwing a ball. We might praise someone for the accuracy or speed of a calculation, just as we might praise someone for the accuracy or speed of a throw. Indeed, many mental activities are subject to blame and praise: plotting revenge, praying silently, making decisions or planning one’s day.

One might object that actions that are not observable by others cannot be the basis for praise or blame; and that when we praise or blame a person for a mental action we are actually praising or blaming them for the overt action or bodily movement that causes us to impute the mental action. For instance, we aren’t really praising Tanya for the quickness
and accuracy of her calculation, but with how soon after looking at the restaurant bill she stated that she owed eighteen dollars, and the accuracy of her statement — this was, in fact, how much she owed. As appealing as this might initially seem, there are numerous problems with this analysis.

Whether or not we can observe an act has little bearing on whether or not we blame or praise someone for it. We censure people for acts that we attribute to them, that is, we praise or blame people for acts that we believe they committed. Often we form beliefs about overt actions based on testimony and indirect evidence. If we see a window broken by a ball, and we see a child with a baseball bat in his/her hand, we might draw the conclusion that the child broke the window by batting a ball at it, despite the fact that we did not witness the act. Or, if a neighbor tells us that he saw Tanya throw the ball towards the window, or even just playing baseball in the yard, we might conclude that Tanya broke the window. The broken window and the testimony suggest that an act of breaking the window took place, but, it is not conclusive evidence since the window could have been broken without having been broken by anyone, let alone by Tanya. Yet, if we believe, rightly or wrongly, that Tanya broke the window, and believe that she did so when acting voluntarily, then we blame her for the act that we did not observe.

Although we cannot observe other peoples' mental actions, we can observe things that provide evidence for mental actions. For instance, if Tanya looks at the restaurant bill, and a few seconds later, announces, "I owe about eighteen dollars," then we conclude that she has calculated her share of the bill, despite the fact that we did not witness the act of calculation. The evidence is not conclusive (she might have guessed), but it suggests that she has performed a mental action — calculating the total. If we believe that she has performed the calculation, we might praise her for the speed with which she did it. When
we do so, we are not praising her for the speed with which she produced an answer, since we would retract the praise if she confessed that she had guessed, or if we discovered that her individual total was already printed on the bill or that she used a pocket calculator. We praise her only because we believe that she did something praiseworthy (calculating quickly and accurately), and the mere speaking of the total does not warrant the praise.

If a man reports that he has thought about stealing from his employer, planning every detail of the theft, then we will probably believe that he has in fact done so, although we have no evidence other than his testimony. Based on this belief, we might blame him for thinking such thoughts: "That's a terrible thing to do." Yet, although we blame him for his thinking about stealing the money, we might praise him for confessing: "It's good that you told me". We might also appraise his act of confession, for being clear, concise or expressive. The mental act and the overt act that provides evidence for the mental act are subject to different kinds of appraisal, suggesting that we blame or praise people for their mental acts, and not just the overt actions that cause us to impute mental acts to them.

Finally, although we cannot observe other people's mental actions, we are aware of our own mental acts. For the most part, we are confident that we know what we are thinking, we are aware that we are silently rehearsing a speech or analyzing a problem, etc. Although we can, of course, be mistaken; we are generally aware of what we are doing in our heads as we are of what we are doing with our bodies. We can, and do, blame or praise ourselves for mental actions. We might feel guilty for plotting revenge or

6 What matters in attributing praise and blame to others is whether or not we believe that they have committed an act. I have claimed that we sometimes form beliefs about both overt and mental actions without observing the act. One might object that overt actions are in principle observable, while mental actions are not. I don't think this is a difference that makes a difference. In either case, we can only observe bodily movements. Even with overt actions, we don't directly observe the acts, we directly observe bodily movements (events) that can be described as actions.
fantasizing about making love to someone, or feel pleased with ourselves for solving a difficult puzzle or memorizing a telephone number.

The fact that we praise and blame people for some mental events shows that we treat those events as voluntary actions; however, it does not help us determine what it is for an action to be voluntary. Aristotle defines a voluntary action as one which is not done by force or through ignorance. Falling, when pushed, is a result of force and is thus, involuntary. Poisoning a dog, by giving it food which was poisoned, is involuntary if the person is ignorant of the fact that the food was poisoned. Yet, although the poisoning of the dog was involuntary, the feeding of the dog was voluntary. Acts done through ignorance are voluntary acts. It is only under descriptions that name particulars about which the actor is pertinently ignorant that they can be considered involuntary. For present purposes, only involuntariness as a result of force helps to capture the distinction between acts and other events.

Things done by force are generally things that the person did not choose or otherwise cause, and thus, cannot be attributed to the person. Many mental events seem to be involuntary in that they are not chosen; however, it doesn't initially appear that they are done by force. Consider mental events of sense perception, like smelling, hearing or feeling cold. In ordinary circumstances, we don't choose to smell an odor, hear thunder or feel cold. These things just seem to happen under certain conditions, for instance, being awake and in the presence of perfume, thunder or cold. We can't just stop smelling perfume, hearing thunder or feeling cold, the way we can stop walking or stop calculating an answer.

We might say that occurrences of sense perception are involuntary because they have physical causes external to us, and that these causes force us to have the sensory
experiences that we have. Minute particles enter our nostrils when we breathe, hit our olfactory cells which respond accordingly, resulting in our smelling the odor. Sound waves hit our ear drums, which respond accordingly, resulting in our hearing. Sensory perception is caused by objects (including photons and sound waves) coming into contact with our sense organs that result in sensory experience.

One might object that sensory experience is voluntary in that we can choose whether or not to open our eyes; we can prevent our hearing by covering our ears; we choose what we put in our mouths. We have some control over what we hear, taste and feel, and even more control over what we see. However, the kind of control that we have over our sensory experience is unlike the kind of control we have over most of our bodily movements. We move our arm muscles just by moving them, and stop just by stopping. We calculate answers just by calculating them, and stop, just by stopping. In contrast, hearing just happens – we hear thunder when it roars, whether or not we are actively listening for it. We must actively do something else to prevent smelling, hearing or feeling cold, for instance, holding our breath, covering our ears or putting on a sweater. Similarly, our heart muscles just move and we can't just stop them, although we can do other things that result in their stopping, for instance, swallowing certain drugs.

Once tripped or hit by a sufficiently large and fast body, we can't help but fall. Yet, we can avoid being tripped or hit if we see it coming and if we can prevent our fall by doing something else, such as grabbing hold of something or adjusting our center of gravity. The degree of control we have over whether or not we fall depends on the amount of experience we have with the things that cause us to fall, the uniformity of the causes, the consistency with which we detect them and our ability to adjust our behavior in order to prevent the potential cause from causing a fall. A toddler falls often, but as he gains
experience with the things that make him fall, and gains more experience with how to adjust his own behavior so as to avoid falling in those circumstances, he becomes a more reliable walker. Indeed, we have a considerable amount of control over whether or not we fall, especially when you consider how many falls we avoid in circumstances that, in the past, resulted in falls — uneven ground, stairways, turning, stepping on something sharp, being pushed, riding a bicycle, walking on ice, etc. The control that we have consists in our ability to sense a potential cause, and to do something in order to prevent our falling, sidestep, catch hold of something or adjust our center of gravity. Yet, despite the control we have over whether or not we fall, falling itself is, for the most part, taken to be involuntary. We blame someone for their fall only when we think that they should have seen it coming and could have done something specific to prevent it. "You knew the floor was wet and slippery; you should have slowed down."

The control we have over our sense perception is of the same type. Once our eyes are hit by photons, we can't help but see; but we have a lot of control over whether and what we see. We can close our eyes, turn off the light, turn away, leave the room, avoid places, etc. As ways of controlling our sight, these are things we learn to do through experience. Unlike falling, seeing is something that we generally desire, so we also learn to control it positively, focusing our eyes, turning our heads, turning on the light, etc. Yet seeing itself is not voluntary the way muscle movements are, including movements of the eye muscles. Seeing happens. Although we blame people for looking at things that don't concern them, we don't blame them for seeing. As many a child has protested when caught prying — I wasn't listening, honest... I just couldn't help hearing.

Sensory perception itself is involuntary, although many acts that involve sense perception, like looking or listening, are voluntary. But even in these instances, when
voluntary actions result in perception, the perception itself is not voluntary. One does not have a choice whether or not to sense, but a choice whether to obstruct, permit or direct the sensation by doing something else — breathing, opening one's eyes, moving one's eye muscles, putting food in one's mouth, etc. These actions are not identical to perception since they do not always result in it.

Many acts that usually result in obstructing, permitting or directing sensation are not treated as fully voluntary. They are better understood in terms of what Aristotle calls ‘mixed’ actions, ones that have both voluntary and involuntary components. Aristotle's central examples are actions performed through fear, which, like involuntary actions, are often pardonable. Throwing goods overboard in a storm is mixed: voluntary since you make a choice, but like an involuntary action in that we are forced to do it by circumstances which are such that no sensible person could do otherwise. Aristotle's account of mixed actions is similar to defenses used in criminal law, such as necessity and duress, that invoke the ‘ordinary person’ or ‘reasonable man’ standard. By this principle, we pardon those who perform what would otherwise be a criminal activity if the act is performed in circumstances so extreme that an ordinary or reasonable person would do the same. We invoke something similar to an ordinary person standard when making everyday moral judgments. We excuse people for what would otherwise be blameworthy behavior if they are motivated by emotions or reasons that are deemed to be natural under the circumstances and which are strong enough that few would resist. For instance, we might censure someone for drinking a beer quickly, yet excuse them on the grounds that they were extremely hot and thirsty. When used, as they often are, to express forgiveness or pardon, the idea that natural emotions can ‘force’ people to behave in an otherwise blameworthy ways, is captured by such expressions as, "I would have done the same
thing," "Boys will be boys," "That's just human nature" and even "Everybody else does it". Some mental actions are pardoned on the same grounds. For instance, if Sarah confesses that she has thought about committing adultery, one's first response may be to reproach her for such thoughts. Yet, if Sarah's husband is known to be an adulterer or if the man in question is overwhelmingly attractive and has made himself available, we might pardon the thought, "Anyone in your situation would think the same thing" (although we wouldn't necessarily pardon the ensuing overt action).

Many acts resulting in sense perception are mixed in the same way. We have a choice as to whether or not we open our eyes (resulting in sight), put food in our mouths (resulting in taste), and even inhale (resulting in smell). Yet, in most circumstances, the desire to see, put food in our mouths and breath, is so natural and compelling that anyone else in the same situation would do the same thing. Although, strictly speaking, we have choice, we don't generally think of opening our eyes, looking around or eating as things we do by choice — we just do them. Such actions, like getting out of bed and getting dressed, are so commonplace that they are not usually subject to praise and blame. They are exempted from moral judgment for much the same reason we pardon people when they are motivated by compelling emotions: these acts, although voluntary, are natural.

We also treat some non-sensory mental phenomena as events rather than actions. We generally do not hold people responsible for images that keep coming back to them or for tunes that keep running through their minds. More often than not, we pity them, especially if the images are ones of traumatic experiences or if the tunes are old Disco hits. We don't usually hold people responsible for emotional responses or for thoughts that 'suddenly strike them'; although we sometimes explain their occurrences by citing triggers or past experiences that suggest complex, and often opaque, psychological explanations.
For instance, the fact that Jordan nearly drowned can be used to explain his feeling of fear when he watches his children swimming. The fact that Laura heard something on the news about an author can be used to explain why it suddenly occurred to her that her library books were overdue.

We treat these mental phenomena as involuntary events; and they too can be shown to be the result of a kind of 'force'. However, in the previous discussions of forced events the cause of the event was external. In these cases of involuntary mental events the causes are predominantly internal, although external events or 'triggers' may play a role in our causal explanations. Aristotle conceives of voluntary actions as ones in which the moving principle which sets the body in motion is internal to the actor. Forced movements are involuntary because the moving principle is external. Yet such an account fails for involuntary events like muscular twitches or tics, and feeling hungry or thirsty, in which the moving principle is undoubtedly internal to the body (although it may be considered external relative to some inner 'true' agent). A more complete characterization of the internal is needed, and I will treat topic in more detail in chapter 5. At this point it suffices to point out that we treat such events as not under our control.

The feeling that thoughts, memories or emotions are forced upon us are captured in expressions such as, 'a thought struck her', 'then it (an idea) hit me', 'he was overcome with fear/desire/anger' and 'it suddenly came to her'. Other expressions portray us as unable to control these mental phenomena, 'the memory stuck with her', 'it kept coming back to him', 'she couldn't rid herself of the notion/feeling', 'he couldn't help but feel/think ....' That some of our involuntary thoughts, images and emotions have internal causes is central to our (incomplete) understanding of dreams, as well as many forms of mental
illness. Similarly, we explain some involuntary overt behaviors by citing internal causes, for instance muscular twitches and ticks, automatism and sleepwalking.

The isomorphism between mental and overt actions extends to so-called 'acts of omission' and the closely related acts of 'standing fast'. We don't just hold people responsible for voluntary actions, but also for voluntary inaction or failure to act. We hold people responsible for acts of omission when we believe: 1) that they were able to perform an action that they failed to perform; 2) that they should have performed an overt action that they failed to perform; and 3) that they knew, or should have known, to perform it under those circumstances. For instance, we hold a driver responsible for failing to signal a turn because we believe that the driver was able to signal, that the driver ought to have signaled, and that the driver knew, or should have known, of that obligation. If it turns out that any of these three conditions are not satisfied, then we excuse the driver, unless we also hold him/her responsible for the conditions not being satisfied. Although a failure to act can hardly be called an overt action, when we hold people responsible for 'acts' of omission we usually hold them responsible for omitting overt actions. However, we also hold people responsible for omitting mental actions; perhaps the most common example is culpable forgetting. I will later argue that remembering is often (although not always) a mental action; at this point it is sufficient to point out that we often treat it as one. "That's an important document, you should have remembered where you put it." "You should have remembered your wife's birthday." In these cases, we believe 1) that the person was able to remember; 2) that the person ought to have remembered but failed to; and 3) that they knew or should have known that they ought to remember. If this doesn't seem readily apparent, consider the following. We would not blame someone who is not able to remember things consistently, for instance, someone known to have Alzheimer's Disease.
Nor would we blame someone who could not be expected to know that they should remember. For instance, we would not blame someone for forgetting his wife’s birthday if he is known to be from a culture in which birthdays are not celebrated, nor would we blame someone for forgetting where she put an important document if she did not know, and couldn’t be expected to know, that the document was important, unless we think that people should remember where they put everything, regardless of importance!

We hold people responsible for acts of ‘standing fast’ or ‘sitting tight’ if we believe that they are under pressure to do something, but voluntarily resist doing it. For instance, in the most literal example of ‘standing fast’, a person resists a force that is pushing him/her, maintaining an upright position. Although the person is not moving, and hence there is no overt action; it is like overt action in that it requires a voluntary contraction of muscles, an exertion of force. We also use such expressions when the ‘force’ being resisted is, as it was for ‘mixed’ actions, a natural and compelling circumstance that would cause most ordinary people to take action. For instance, a person stands fast who, in the face of great danger, resists the natural impulse to flee. Standing fast and sitting tight are also used metaphorically to explain mental ‘acts’. For instance, when one resists changing one’s mind when others are disappointed with the decision, when one resists reconsidering a position or belief in the face of compelling evidence to the contrary, or when one resists the temptation to think lustful thoughts about an attractive person. The isomorphism between mental and overt acts of standing fast is, perhaps, most apparent in such expressions as, "She didn’t budge from her position," which describes an effort to resist a compelling force, which might be physical or mental.

The last pertinent similarity between overt and mental acts is that we give the same types of explanations for both. The brain is often considered the immediate cause for both
types of events. Although neuroscience is still very much a developing science, and we
don't know exactly how different types of actions and other behaviors are caused, we
generally accept that my both throwing the ball and calculating the answer involve neural
activity which is immediately responsible for the behavior, and that such neural activity is
itself caused by neural activity. Other forms of causal explanations also apply to both
mental and overt acts, for instance, belief and desire explanations. Often the expected
answer to "Why did you do that?" is a belief, a desire, or both, whether or not 'that' action is
mental or overt. Consider the questions, "Why did you take that slice of cake?" and "Why
did you calculate the answer?" In both cases, the explanation would involve a desire (I
wanted to eat the cake, or I wanted to know the answer) and a belief that performing such
an action would result in getting the cake or the answer.

Our taxonomy of mental events is isomorphic to our taxonomy of bodily
movements: some are involuntary and some are voluntary. Of those that are voluntary,
some escape moral sanction because they are motivated by compelling, but natural,
circumstances. We hold people responsible for omissions of both mental and overt acts,
and for acts of standing fast, whether they resist with muscle or with mind. This taxonomy
of mental and bodily events reflects a systematic categorization based on notions of
compulsion, choice and control. It is on the basis of this categorization that we regard
some events as actions of agents, and hold those agents responsible for their actions, and
do so whether the events are overt or mental acts.

One Theory of Human Behavior

The isomorphism between the taxonomy of mental and overt events, behaviors and
actions suggests that a single theory of human behavior should be able to account for both
mental and overt behaviors. After all, since mental and overt behaviors have such strikingly similar properties, and since they appear to have the same immediate cause (the brain), a reasonable hypothesis is that they are, in fact, produced in the same way. Such a theory should account for mental behaviors (those activities that are identified with the mind: thinking, imagining, calculating, holding of propositional attitudes, etc.) on the same basis as it accounts for overt behaviors (those activities that are identified with the body: walking, talking, promising, throwing balls, firing guns, etc.) This can provide a new picture of the mind and of the relation between 'mind' and 'body'; a physicalist picture that neither eliminates mental phenomena nor reduces them to neuro-physical phenomena.

In this essay, I advance a theory in which behavior is a single general kind, of which the mental and the overt are merely two interestingly distinct forms. Many behaviors, whether mental or overt, are learned, acquired or conventional behaviors that depend on prior events in which a behaver learns from sensory experience of the world and him/herself. From this interaction with the world, our behaviors come to represent the world around us, and can be interpreted as meaningful within the context of shared experience or a shared environment. Furthermore, since other people form the most important part of our environment, nearly all behaviors have strong social components and consequences. Some behaviors, those which we might call practices, are so dependent on a social environment that they are only meaningful within the social context in which they arise.

In the following chapters I outline a theory of human behavior with an emphasis on understanding mental behaviors in the same way that we understand overt behaviors. In keeping with my behaviorist starting point, learning and the development of habits are of central importance. Learning is, on my account, primarily due to our forming neural
'representations' as a result of our conscious interaction with the world. Thus, I begin in chapters two and three with a discussion of representation in general, and more specifically, neural representation. I present an account of representation in which the human being can be seen as a medium of representation: the learner is an active medium of representation, developing stable behavioral habits in accordance with complex neuro-representations which serve to organize her perceptual experience.

The theory of representation I advance is compatible with and, to some extent, inspired by connectionist models of the brain. However, it is also a return to more traditional theories of the mind, like those of Aristotle, the Stoics and Descartes, in so far as I take natural representation and realist art as models for neural representation, and in so far as I argue for a kind of semantic internalism. Yet in contrast to other 'picture' or 'impression' theories of mental or neural representation, I urge a conception of the human being as a medium of representation, thereby emphasizing our sensory interaction with the world.

In the fourth chapter, I extend the discussion of representation to language use, or that part of it which I take to be linguistic representation. Our ability to mean things with words stems from our capacity to represent the world and to utilize these representations in the production of meaningful and interpretable behaviors. But, although linguistic behaviors depend on our capacity to represent the world, they are much more social or conventional than other forms of representation. Thus, language functions at the interface between the private, mental realm and the public or social environment in which we acquire and use it. Linguistic representation has two faces, one which is semantically internalist (our meanings are 'in our heads', personal or idiosyncratic) and one which is semantically externalist (we learn to use language, to mean things with our words, because we observe
other people doing things with their words — things that we also learn to do with our words). We mean things in many different ways, in different contexts. This contextual dependency itself suggests that we primarily function with an externalist semantics — a familiar viewpoint supported by many twentieth-century philosophers, such as Wittgenstein, Quine, Putnam, and Davidson. This external semantics is, on my account, made possible by an underlying internalist semantics which is closely tied to the theory of representation advanced in chapters two and three.

The theory of representation, linguistic and non-linguistic, provides a working model of the relation between people’s brains, their behavior and the world which they sense and from which they learn. I present only a sketch of a theory of representation and language use; it is, however, sufficient for the larger project of developing a theory of mental and overt behavior. Although this larger project does not rest on this particular theory of representation it does require a theory similar to it in important ways:

1) The relation via the senses between a person and his/her environment is central to the production and understanding of behavior.

2) Representation is understood as a set of systematic alterations to a medium of representation, rather than resemblance, but is not understood in terms of signs, or straightforward covariance between representational structures and external objects or causes.

3) Representational structures can be utilized independently of being interpreted (for instance, in the production of new representations or behaviors).

4) Like other meaningful behaviors, linguistic behaviors are dependent on non-linguistic but representational, or meaningful structures.

On such a theory of representation (which I call the medium theory), behaviors are meaningful in so far as they utilize representational structures, and thus the behaviors are both representational and interpretable. Their meaning springs from the past relations with the world which shape those representational structures. When a person’s behavior is
produced by structures that have been produced by that person's experiences, then those behaviors can be interpreted as meaningful in the context of such experiences.

The medium theory of representation provides a framework in which to understand meaningful behaviors, and thus, to distinguish between different types of behaviors, whether intentional, deliberate, automatic, habitual, and whether they are overt or mental behaviors. In the second part of this work, I turn my attention specifically to mental behaviors, actions and practices. Mental behaviors, on this theory, are representational and, therefore, interpretable since they are produced by representational neural structures in the same ways that overt behaviors are. However, mental behaviors differ from overt behaviors in that they involve the suppression (to anticipate a technical term which I introduce in chapter 5) of overt bodily movement.

In the fifth chapter I address the general nature of behaviors and actions, and more specifically internalization of overt behaviors through the suppression of overt bodily movements. The discussion of suppression centers on two of the most important mental behaviors: desiring and thinking in words. In the course of the discussion, I distinguish, on the one hand, between movement and behavior, and on the other, between behavior and action (or intentional behavior). The resulting account of rudimentary action (object-oriented behaviors following an interval of suppression) leads to the discussion in the sixth chapter of more complex forms of action, including 'situated' behaviors, speech acts and deliberate action.

In the seventh chapter, I discuss the different ways in which we do things 'in our heads' and different consequences of these mental actions, the ways in which we perform them and the circumstances in which we perform them. The consequences that are of greatest interest concern our conceptions of agency and of the mind. Here it is of central importance that we learn to internalize overt behaviors within the context of the overt
practices and activities that constitute a society. This leads to, in the eighth and final chapter, a characterization of the things we do 'in our heads' as constituting a mental life. Just as our life is comprised of all that we do and all that happens to us, our mental life is comprised of all that we do and all that happens in our conscious experience. Thus, this approach to human behavior yields a picture, not of the mind, but of a mental life which can take on many forms. The dominant forms of our mental life are best understood within the context of the overt activities and practices which constitute the social context in which a person develops.
PART I

THE MEDIUM THEORY OF REPRESENTATION
Chapter 2

PEOPLE AND OTHER MEDIA

For centuries the mind and brain have been compared to various media of representation, reflecting the sensible world in much the same way that many works of art represent their subjects. The mind or brain is a medium in which objects are impressed or painted, as seals are impressed in wax, faces cast in plaster, scenes painted on canvas or drawn on sketch pads, moments captured on photographic film, or action recorded on video tape or reproduced on stage.¹ The success of these metaphors, I will argue, rests on a causal and historical relationship between represented objects and the medium in which they are represented. By considering the medium-dependent nature of representation, we can see how human minds, brains and behaviors can be understood as representational, and thus, as interpretable by ourselves and by others.

Brains, minds, wax tablets, film, painted canvas and plaster casts have a similar representational capacity that is best understood in terms of the causal and historical relations between the medium and the object represented. This capacity derives from the presence in them of what I call 'history-laden structures'. A history-laden structure represents an object in a medium such that the object was causally responsible for the structure and such that the medium ensures that key features of the structure in the medium co-vary with key features of the object. In interpretation, we must take into account both the medium and the structure in the medium. By considering key features of the structure, and the ways in which the medium ensures that those features co-vary with

¹ The list of people who have used these metaphors for the mind, the soul or the brain spans the centuries — Plato, Aristotle, many Stoics, Augustine, Descartes, Locke, Hume, some contemporary philosophers and also psychologists such as Stephen Kosslyn (1994).
features of the object, we grasp the structure as representing the object, despite differences, possibly dramatic differences, between the structure and the object.

For instance, consider an envelope sealed with a blob of wax. The structure of the blob of hardened wax can be understood as representing a certain signet ring\(^2\) because the signet ring was causally responsible for the shape of the wax and because the medium of wax ensures that key structural features of the wax co-vary with key structural features of the ring. Consequently, there are two different ways in which the wax structure might fail to represent a signet ring: 1) if a signet ring were in fact not causally responsible for the wax having that shape; and 2) if the medium of wax was not sensitive to signet rings, and thus, did not ensure that key structural features of the wax co-vary with key structural features of the ring. Thus, if the wax had been impressed by something other than the signet ring, it would represent that which impressed it rather than the signet ring.\(^3\)

Likewise, the wax blob would fail to represent the signet ring if the wax itself was an insensitive medium of representation. For instance, a medium sensitive to signet rings is soft at the time of imprinting, then quickly hardens to retain the impression: sealing wax is

---

\(^2\) Of course, it doesn't just represent a certain signet ring. We use seals for all sorts of purposes. Yet, some of those purposes, for instance, identifying or verifying the source of the letter, rely on the blob's capacity to represent that signet ring to those who examine the letter.

\(^3\) There are a number of indirect ways in which a signet ring, or other object, could play a causal role in shaping the wax and thus be indirectly represented, and appear as if it were directly represented by the wax. For instance, if a forger manipulates the wax so that it looks as if it were impressed by the signet ring. Or, if a forger makes a mold from other wax impressions of the signet ring and uses it to impress the wax. In such cases, the original signet ring would need to be mentioned in a full causal account, and thus, plays an indirect causal role. In such cases, the wax does not directly represent the ring, since the ring is not directly responsible for the wax structure. Yet, despite this, it would be reasonable for a person to interpret the wax as representing the ring (in the usual, direct sense) if he/she were unaware of the role of the forger but had knowledge of the medium (sealing wax) and its capacity to represent things impressed in it. This is why forged seals can function as 'the real thing' while failing to be authentic, but only up until the point at which the history (of an indirect causal role) is discovered, at which point they cease functioning as 'the real thing' and begin to function as 'forgeries'.


a sensitive medium, but, in most contexts, beeswax is too soft and resin is too hard and brittle.$^4$

The represented object, then, is that which is causally responsible for the history-laden structure in the medium. Obviously, the represented object is not necessarily the same as that which we interpret a structure as representing, since when we interpret a representational structure we may not interpret it as representing the very object (or objects) that is (or are) in fact causally responsible for the history-laden structure. When such a structure is interpreted, the interpreter reasons from her observations of the structure, and her knowledge of the medium in its capacity to represent objects. In many cases, a skilled interpreter will be able to infer from the representational structure and her knowledge of the medium to certain features of the represented object. Thus, a skilled interpreter observing the wax seal, and knowledgeable about the capacity of sealing wax to represent objects, can infer from the seal to the represented object — a solid disk with a certain design inscribed in it. These features inferred from knowledge of the medium and observations of the history-laden structure in that medium may be insufficient to identify the unique object that was causally responsible for the history-laden structure. However, the interpreter may have additional information that allows her to further infer that the object of representation is a particular object — for instance, having seen a gold signet ring with that particular design on it on Simon’s finger last month, the interpreter may be able to infer that the seal represents Simon’s ring.

$^4$ Stephen Kosslyn makes a similar point with respect to mental imagery. He writes, “The important implication of this idea [Plato’s comparison of memory to the impression of seals in wax] for present-day investigations of imagery is the notion that we can distinguish between the properties of the medium (the wax tablet) and those of the representation of information (the etchings). No matter what information is stored, it will not change the nature of the medium — wax is wax. The properties of the medium, however, affect all the representations that occur in it. No matter what is being etched, if the wax is unresponsive, the etcher is in trouble.” (Kosslyn 1983:54)
The varied ways in which an interpretation can succeed and fail, or be justified or warranted, depends on the skill of the interpreter (in observing and reasoning) and the kind of knowledge available to the interpreter (knowledge of the medium, the history-laden structure, the context, and socially relevant standards and methods of interpretation). In addition, in many contexts, it may be reasonable to interpret a history-laden structure as standing for something other than the represented object. For instance, in some contexts, a photograph of a crying child may be better interpreted as standing for sadness, than as a representation of the particular child which the photograph actually represents. There are important questions concerning interpretation that I take up elsewhere; but first I want to consider the nature of representation itself. Ultimately, our ability to interpret behaviors, objects and ourselves as meaningful, depends on a natural, causal relationship between objects and a medium. It is this relationship that I investigate in this chapter.  

Before embarking on a more detailed explication of history-laden structures and their capacity to represent, it is important to distinguish representations from signs and symbols. Representations, signs and symbols can all be seen as standing for other objects. I use ‘standing for’ as a general term to capture the shared relation which obtains

---

5 The examples that I use in this chapter tend to be the simpler ones that involve only simple relations between media and objects, in which the history-laden structure is produced through a single event in which the medium responds to the object. Yet, some representational structures result from multiple events, each of which involved the medium’s exposure to an object, or to different objects. In these cases, there would be a complex or multiple cause. Consider time-lapse photography, or other photographic effects produced through multiple exposures. In keeping with contemporary theories of the brain, particularly connectionist accounts, neuro-representations would often (but not always) result from multiple experiences with certain types of objects (colors and shapes, animals, moving objects, spoken or written words, etc.) Representational structures in the brain often have complex or multiple causes, especially neuro-structures that might be said to represent beliefs, language, or concepts, which we generally acquire over time and through many experiences.

In addition, a photograph may represent both the objects that the camera was pointed at and the mood or feeling that the photographer intended the photograph to capture. In such a case, both the objects in the world and the photographer’s intentions are causes of the history-laden structure, and may both be represented by it at one and the same time.
between representations, signs and symbols and their respective objects. I distinguish these three relations (signification, symbolization and representation) in the following way:

**Signification:** A signifies (or is a sign of) B if A regularly occurs with B. For instance, smoke regularly occurs with fire, and this consistent co-occurrence is what I mean by the relation 'signification'. Hence, most people with experience of the co-occurrence of A and B will interpret A as standing for B.

**Symbolization:** A symbolizes (or is a symbol of) B if A frequently occurs with B, but co-occurrence is a result of human causal interaction. For instance, evergreen trees occur with Christmas (in some highly specialized contexts) due to human convention, and thus evergreen trees have come to symbolize Christmas in those contexts. People mark bottles of poison and pirate flags with the skull and crossbones, ∂, and thus, it symbolizes such things as death, a skeleton or danger, and might constitute a warning (of pirates or poison depending on the context). Likewise, due to the scientific conventions, 'c' is a symbol for the speed of light in certain contexts. This frequent co-occurrence in such contexts, due to human causal interaction, is what I mean by the relation 'symbolization'. Hence, most people with experience of the co-occurrence of A and B in such contexts will interpret A as standing for B in those contexts.

**Representation:** A represents B if A is a structure in a medium caused by interaction with B such that the structural features of A co-vary with structural features of B. For instance, if an impression in wax is a structure in the wax caused by interaction with a signet ring, such that structural features of the impression co-vary with structural features of the signet ring, then the wax impression represents the signet ring. Likewise, drawings and photographs can represent people or places, and phonographic records can represent songs or music. Thus, most people who are familiar with the medium of representation, will interpret structural features of A as standing for structural features of B, and will interpret A as representing B.

I take as central to the notion of representation: 1) covariance between structural features of the representation and structural features of the represented object; and 2) an important

---

6 The only difference between signification and symbolization is that in the first case the co-occurrence is natural, and in the second case, it is conventional. Thus, it is sometimes difficult (even impossible) to determine whether or not something is a natural sign or a conventional symbol. For instance, consider that ripe pumpkins, gourds and colored leaves stand for autumn in North America. They are successful symbols for autumn precisely because they are natural signs of autumn. Many bodily expressions are also indeterminate between natural signs and conventional symbols: shaking one's head, shrugging one's shoulders, raising one's eyebrows, etc.

7 I am not going to talk about words or concepts as signs, symbols or representations of objects. I think that the ways in which words stand for objects is quite unusual. The present theory of representation is the basis for a theory of linguistic meaning. Neural representations (which this theory does account for) are best understood as non-verbal tacit beliefs from which overt and mental linguistic behavior can be produced. A full account of linguistic meaning will be given in chapter 4.
dependence on the medium of representation. Unlike representations, signs and symbols neither depend on a medium nor do they involve covariance between structural features of the sign or symbol and their objects. In contrast, signs and symbols display a more direct covariance between the sign or symbol as a whole and the signified or symbolized object as a whole.

Signs and symbols co-occur with the objects signified or symbolized, as fire co-occurs with smoke, chimneys with fireplaces, red octagons with the word 'stop' and certain letters with certain phonemes. But, structural features of the sign or symbol do not usually co-vary with structural features of what it stands for. Of course, some symbols are also representational, at least in origin, for instance, consider the symbol ☎. It stands for telephone, and is used as a symbol for phones because it visually represents a phone, even though most phones don't look like that anymore. (Indeed, contemporary phones may not have a distinctive enough shape to be adequately represented by a simple symbol.) Thus, a symbol or sign may retain representational features, but nevertheless function merely as a symbol or sign, rather than as a representation. It functions as a symbol in that we expect the symbol of the telephone to indicate the presence of a telephone, but we do not expect to draw conclusions about the structure of the phone from the appearance of the symbol.

In contrast, consider a pictorial representation: a map of the United States of America. Shapes on the map co-vary with shapes in the world: for instance, rectangular shaped patches on the map (labeled 'North Dakota', 'South Dakota' or 'Colorado') co-vary with rectangular shaped areas of land (North Dakota, South Dakota, and Colorado); relatively small patches on the map (labeled 'Maryland' and 'Connecticut') co-vary with
relatively small areas of land (Maryland and Connecticut). Likewise relational properties, being to the right or above on a map co-vary with being to the East or the North out in the real world. The map as a whole is a representation of an area in the world, because structural features of the map co-vary with structural features of the world and because it was due to the fact that the world had those structural features that the map was drawn to have those structural features. If the world had been otherwise, so too would the map have been.

The success of a history-laden representation depends on the medium’s capacity for representing. Color film, canvas and paint, paper and pencil crayons can be good media for representing colorful visual objects, but are less successful media for representing sounds, flavors, smells and movement. Audio tape, musical scores and phonographic records are good at representing music or sounds; smells and flavors can be good at representing each other (consider ‘scratch and sniff’ books); and video tape and flip books can be good at representing movement.

The relation between the medium and the type of object represented is complex. Some kinds of media are more easily shaped by certain kinds of objects, are more easily restructured to reflect certain aspects of those objects, or preserve such changes better, and thus are better at representing those kinds of things. For instance, consider running rabbits. As the rabbit runs across a surface, he comes into contact with the surface,

---

8 Relative to Texas, North Carolina, etc.
9 It may be an abuse to say that a smell can represent the flavor of something, mostly because we don’t have an widely shared vocabulary adequate for describing the closely related features of odor and taste. Perhaps the vocabulary of wine tasting comes the closest. I think that smells and flavors can, in some cases, represent, not only each other but other things, although they more often symbolize or signify. Consider how the distinctive smell that is added to otherwise odorless natural gas symbolizes natural gas in our culture (indeed, this is also a good illustration of the relation between conventional symbolization and natural signification. Many people are unaware that this odor is added to natural gas and is not its natural smell, thus, for those people, it may be more accurate to say that it is signifies rather than symbolizes natural gas), or how some food is ‘comfort’ food — certain flavor/texture combinations symbolize maternal attention.
causing changes to the surface at the point of contact. We call such structural changes rabbit tracks. Different surfaces respond differently to rabbits running across them, and thus have different capacities for representing such events. For instance, just because they undergo more (and longer lasting) structural change when rabbits run across them, snow, wet sand, and mud are generally better at representing running rabbits than are grass, dry sand and pebbles, which are nonetheless better than ice and water. The best media for rabbit tracks, and thus the best surfaces for representing running rabbits, are surfaces that are changeable by, or sensitive to, running rabbits and that preserve those changes (as mud does, but water doesn’t).

Rabbit tracks represent many aspects of the rabbit, in the sense that structural features of the rabbit tracks vary with aspects of the running rabbit. But, exactly what features will be represented depends on the medium. For instance, consider rabbit tracks made in near-zero temperature packing snow. The shape and size of the individual paw prints are due to the shape of the rabbit’s feet. This can be expressed as a counterfactual: if the rabbit had had different shaped feet, the shape of the prints would also have been different. The spacing between the tracks is due to the spacing between the rabbit’s legs, and also to how fast the rabbit was running. Had the rabbit been bigger, or had the rabbit been faster, the tracks would have been farther apart. The depth of the depressions represents the momentum of the rabbit. Had the rabbit been heavier, or had it been jumping higher and faster, the rabbit tracks would have been deeper. Yet, the color, gender or age of the rabbit, the shape of the rabbit’s ears — none of these things are represented in the rabbit tracks, except in so far as they co-vary with the things that the tracks in the snow can represent, for instance, young rabbits generally have smaller prints that are closer together. The age of the rabbit is not represented in the tracks, although
the size of the rabbit is; and thus, one might be able to infer the age of the rabbit from other features of the rabbit that are in fact represented in the tracks. But, the medium itself determines which features of a type of object can be directly, or indirectly, represented by structures within it.

The fact that the medium of representation determines which structural features are representational is reflected in acts of interpretation. In the example of the signet ring, the wax is sensitive to surface texture of hard objects, but not to color. When interpreting such structures, we do not draw conclusions from the representation in the wax about the color of the signet ring. Likewise, if we have experience with black and white photographs as media of representation, then we do not draw the conclusion that represented objects were shades of gray from the fact that the images in the photographs are black, white and gray. In order to interpret photographs and impressions as representations, one must usually have some experience with the medium of representation. For this reason, young children do not, as a rule, immediately recognize maps, pictures, photographs and videos as representations.

The capacity of a history-laden structure to represent objects is independent of interpretation, although, of course, referring to such a structure as a representation of an object or even as co-varying with an object entails the possibility of observation, and thus, interpretation. A structure is representational if it meets the following sufficient conditions: 1) the structure exists in a medium which is sensitive to certain kinds of objects; 2) the structure is causally connected to an object (or objects) of that type. If these conditions are met then it will be possible for a knowledgeable observer to interpret the structure as a representation of that which it is historically or causally connected to. But, interpretation itself is not a necessary condition for representation. In contrast, every act of
interpretation involves treating an object as a history-laden structure which represents that to which it is directly or indirectly causally connected.

When we interpret a representational structure, we utilize our beliefs about the medium as a medium of representation and our observations of the structures in the medium. By looking at a photograph, and knowing something about photography as a medium for representing the world, we can interpret the photographic image as representing the world. We could not do this so successfully if the medium of photography were not sensitive to the kinds of objects involved. When the medium ensures covariance between structural features of the structure in the medium and structural features of that which caused the structure, then an interpretation based on the history-ladenness of the representation is theoretically possible. Thus, if a photographic image was 'accidentally created'\(^\text{10}\) on dark wallpaper, then it represents what it represents even in the absence of an interpretation or an interpreter. If the medium is sensitive to the object in this way, then there is a nearby possible world in which an interpreter capable of interpreting such structures exists whether or not such an interpreter could exist or does exist in our world. Thus, the capacity of a history-laden structure to represent is independent of an interpretation or interpreter, but ensures that an interpretation is possible, provided certain other conditions are met (for instance, a sentient and interested observer).

The medium of representation determines which key structural features of the representation will co-vary with key structural features of the represented object in a non-arbitrary fashion. Resemblance is a generally recognizable result of covariance, but covariance does not always produce resemblance. History-laden structures in some media

\(^{10}\) If a room has dark wall paper, then if there is only a pin prick of light coming from outside, then the effect of the camera obscura could cause fading of the wallpaper, such that a 'photographic' image could be created by accident.
will produce a very strong resemblance between structural features of the representation and structural features of the object. But there will also be failures, or ways in which the representational structure fails to resemble the object. In Languages of Art, Nelson Goodman takes such failures of resemblance as marks in favor of a denotational theory of representation against resemblance theories of representation, like that held by Ernest Gombrich. Goodman argues that because of differences in perspective (among other things) photographs and representational pictures must fail to fully resemble that which they represent. He writes, "the behavior of light sanctions neither our usual nor any other way of rendering space; and perspective provides no absolute or independent standard of fidelity." (Goodman 1968: 19) Although I am in agreement with Goodman that a theory of representation cast primarily in terms of resemblance is inadequate, his argument against such theories fails to consider the role of the medium.

Gombrich, when responding to Goodman, rightly directs attention to our trust in our familiarity with the photographic medium:

Some time ago, while vacationing in the Austrian Alps, my eyes fell on a picture postcard that showed a splendid view of the upper reaches of the Inn valley towards the Swiss frontier. It looked as if it would be worth seeking out the place from which the photograph had been taken. Since I neither doubted its fidelity, nor was unaware of the behavior of light, I took a map and a ruler and attempted to plot the vantage point by the direction of the river and the parts of the vista visible on the picture. We then set out with the card and the map and were rewarded by the beautiful view. Not quite, though: we saw the valley along the same axis, but the card proved that certain distant mountains had also been visible to the camera which showed less on our horizon; the photographer must have stood a little higher, at a spot where it was too late for us still to climb. But this discovery, of course, did not shake our confidence in the method of plotting the spot. On the contrary, it was due to that method. Whatever we may mean by "ways of rendering space" we all rely on photographs or topographical views to tell us what objects in space can be seen from a given point. Clearly, we also know that the relationship between maps and such pictures is reversible. Just as a picture can be located on a map, so a map allows us to construct a picture. (Gombrich 1972: 129-130)
For history-laden representational structures, both the failures and the successes in resemblance will be predictable and accountable in terms of the difference between the medium of representation and the material of the original object, and the difference between our relation to the representation and our (possible) relation to the original objects of representation. For instance, photos are two-dimensional, and the images in the photograph are all (roughly) the same distance from our eyes. In contrast, the objects of which the photographs were taken are often three dimensional and at different distances from our eyes. Likewise, the effects of light on photographic film through shutters are dissimilar to the effects of light on us through our eyes. These differences account for some of the failures of similarity between the images in the photograph and the objects of which the photograph was taken. As Gombrich reminds us, when we have adequate knowledge of the medium of photography, we can take advantage of these regular differences and use them in reconstructing the original objects of which the picture was taken.

Consider how it is that a photograph of a person represents that person. Most people have an idea of how photographs get made: a camera with film in it is directed at a scene, when the button is pressed something involving light happens, and then, in a dark room, pictures are made from the exposed film. Thus, if everything works out, the photograph is a representation of what the photographer took the photograph of. By a common standard of evaluation for snapshots and identification photos, a photograph is a good representation of a person in so far as the photographic image of the person appears visually similar (for a typical observer) to the person of whom it was taken. Thus, for most amateurs, color photography has an advantage over black and white photography in so far as the photographs are better visual representations of the subjects.
Sometimes, for various reasons a photograph can succeed in being a photograph of a person while failing to be a good likeness of the person. Thus, a sufficiently bad photograph of Mary can be a very poor representation of her. Perhaps a shadow made her look as if she had a black eye, and the lighting conditions made it appear as if she had red eyes and was light skinned, etc. But, there is an important sense in which the bad photograph of Mary remains a representation of Mary rather than a representation of Johannes or Cordelia (which it can be mistaken for), since, after all, the photographer pointed the camera at Mary and not at Johannes or Cordelia. Photographs can also succeed in being very good representations of a person and yet at the same time bear a similar likeness to another person. If Cole and Colin are identical twins, then a photograph of either twin may bear the same likeness to either twin. Yet, any such photograph is causally connected to either one twin or the other, and thus the actual represented object is the twin to which it is causally connected (whether or not we can determine which twin is actually represented from looking at the photo).

We can take a plaster cast of a person’s face, and from that cast make an image of the face. Depending on the skill of the caster, the likeness will be better or worse. Yet, as with the photograph, no matter how close a likeness the finished product bears to the person, if the finished product is a representation at all, it is a representation of the person who was cast in plaster. Similarly, footprints in the sand represent the feet that made them. Thus, when Robinson Crusoe saw Friday’s footprints on the beach, he saw them as a representation of a man: his knowledge of wet sand as a medium of representation together with his observations of the structural features of those prints allowed him to infer to certain structural features of the object of the representation, i.e., that he was a man and quite recently present, since the prints were shaped like large human feet and wet sand is
a very temporary medium of representation. Likewise we reason from the structural features of certain indentations in old river beds to the structural features of dinosaurs, because we recognize that these structures are history-laden and know something of the capacity of the medium (fossilized mud) to represent certain kinds of objects.

When we consider examples such as footprints or animal tracks, it becomes clear why a resemblance theory of representation is inadequate, even if one takes the medium into consideration. To an experienced tracker, rabbit tracks can represent the approximate speed of the rabbit, and the approximate size and weight of the rabbit, and even the activities of the rabbit. Yet, the only visual resemblance is between the shape of the rabbit prints and the two dimensional shape of the rabbit paws, and between the spacing of the prints and the spacing of the paws when they touched the ground. There is even less resemblance between digital formats (such as CDs) and the sounds or visual objects that they represent, or between exposed but undeveloped film and the objects that it represents.11

One might argue that for this reason the tracks, CDs or undeveloped film signify or symbolize, rather than represent. The failure of full resemblance, after all, caused Goodman to opt for a denotational theory of representation (Goodman 1968). Yet, signification and symbolization also seem inadequate since they fail to capture the role of the medium of representation. Consider how ‘signify’ is used in the sense of a ‘natural sign’. Normally in the case of signs, there is a simple one to one relationship, without a role being played by a medium. Smoke signifies fire directly, and certain kinds of ringing sounds signify telephones. In the case of rabbit tracks, the relation is much more complex, as it is the nature of the medium (the snow) and its responses to rabbits that make many

11 I will argue later that these ought also to be considered representations.
distinct structural features of the tracks interpretable as representing corresponding
structural features of the rabbit. Rabbits make tracks in the snow by modifying the
structure of the snow in ways that are preserved through time. In order to interpret tracks
as representations of rabbits we must know something of the nature of the medium. In the
case of signification the medium does not play this sort of role.¹²

The medium also fails to play this type of role in symbolization, despite the fact that
a medium is often involved, and despite the fact that the symbol itself may be
representational (or pictorial). Consider pictorial symbols such as ☐ or ❌. Both are
pictorial and both require a medium. But, knowledge of the medium does not determine
the ways in which structural features of the symbol co-vary with structural features of the
object symbolized. Knowing that the symbol is drawn on paper does not help us
understand what the object it symbolizes is like — it does not help us understand how
structural features of the object it symbolizes are similar to or covariant with structural
features of the representation. At best, the medium helps determine context. For instance,
when the skull and crossbones is drawn on a piece of paper (attached to a jar) it likely
symbolizes the presence of poison, and when it is on a white flag (at the top of a ship) it
likely symbolizes pirates. Consider how a written word is a symbol for a spoken word¹³.

¹² Representations can also function as signs and/or symbols. Animal paw prints are quite popular
symbols in zoos (for marking paths), on environmental advocacy materials, on cards and wrapping
paper, as well as in children’s games and toys. There is not always a clear boundary between
representation and signification or symbolization, and in many cases, whether a structure functions
as a representation, a sign or a symbol depends on the context and the skills and knowledge of the
observer. My own tracking skills are quite limited, and thus I see animal footprints in the snow as
signs of animals rather than as representations. Yet, despite the fuzzy boundaries, representations
can be distinguished from signs and symbols as medium dependent and involving covariance
between structural features and not merely covariance between occurrences.

¹³ The relation between written symbols and a word, or letters with phonemes, is a symbolization
relation: the written symbols symbolize the word. This is completely independent from the way in
which a word or concept stands for objects or ideas, which I am not concerned with here. As noted
earlier, this theory of representation does not apply to linguistic representation, but will be used as a
basis for a theory of linguistic meaning.
Although media are involved, the media do not play any role in determining what is symbolized or what the nature of the symbolized object is. 'Antelope', 'ANTELope', '__Antelope__' and 'antelope' all signify the same spoken word (in most contexts) and they would continue to do so if written on different materials or carved in a block of wood. But, our knowledge of the medium in which the word 'antelope' is written does not contribute to our knowledge of the correct pronunciation of the word or other structural features of the spoken word. Our knowledge of the medium does not determine which structural features of the spoken word should be inferred from the structural features of the written word.

Only for history-laden structures does knowledge of the media allow us to reason from structural features of the representation to features of the represented. Rabbit tracks share with photographs, video, pictures and prints in wax an important dependence on the medium of representation. For the most part, one can reason from the representation in that medium to the represented object only if one is familiar with the type of medium involved. From knowledge of the medium, one can often predict which structural features of the representation will be representative of which structural features of the object, as well as to what degree and in what ways the original object will resemble the representation of that object. Black and white photographs don't resemble their objects with respect to color, but anyone familiar with black and white photography can reasonably interpret objects in black and white photographs as colorful. Images in photographs don't resemble their objects with respect to size, but if you are familiar with photography, photographs can be quite useful representations of the sizes of things.

There are interpretational advantages to history-laden representational structures that are also due to the role of the medium. The first time a person has experience with a non-representational sign or symbol, he or she cannot interpret it as representing any
particular object just from consideration of the symbol or sign itself, (although the context may indicate that it is a symbol or sign, and, in some cases, provide clues as to the nature of the object). In general, it is only after I have experienced the co-occurrence, learnt a convention or heard the stipulation, that I can understand the sign or symbol as standing for that with which it co-occurs, whether naturally, conventionally or by stipulation. The first time one sees an unfamiliar symbol, even if it is pictorial, and in a familiar medium, one would be foolish to reason from it to structural features of the object symbolized. For instance, consider the symbol  המשנה. When you see it for the first time, even if you recognize it as a symbol, it would be foolish to reason from the structural features of the symbol to structural features of the object that it stands for. Even if you recognize it as a window in motion, the most one ought to do is reason that whatever it symbolizes has some connection with windows in motion, and perhaps make an educated guess that it is a symbol for the computer program, Microsoft Windows®.

In contrast, provided one has experience with a medium in its representational capacity, then one can interpret a completely novel history-laden structure in that medium as a representation of the object that caused the structure. The very first time one sees an unfamiliar representational structure in a familiar medium, one can reason from structural features of the representation to structural features of the represented object according to our knowledge of the medium. At the first occurrence of an unusual set of prints in a riverbed, I can begin to reason from the structural features of that representation to features of the represented object. I may be wrong and misinterpret what it represents. But, often, such reasoning is quite successful, and the mistakes or misinterpretations are easily understandable in terms of over- or underestimating the capacity of the medium to represent (the foot shape is misrepresented because there is some distortion when the wet
clay in which the prints were made dries), or making unwarranted inferences from structural features of the represented object to the type of object it was (the prints may correctly represent the animal as very heavy, and three-toed, but this does not warrant my inference that the heavy, three-toed animal was a Tyrannosaurus Rex; it may in fact have been a gigantic prehistoric chicken). We only interpret such history-laden structures by reasoning from the structural features and our knowledge of the medium because the method works so well in general.

The role of the medium is of utmost importance in 'direct' representation, for instance, wax, plaster casts, snow, photographs, audio tapes and video tapes. But media that involve a more active human or intentional role can function as representational media in the same way. If an artist is skilled and consistent, and aims to represent the world as it appears to us then we can interpret the painting pretty much as we interpret a photograph. The goal of producing a realistic representation of the world has often been a laudable aim in art, and was of utmost importance in the Western tradition. Gombrich, following Constable, compares art in the Western tradition to a science. The comparison is apt in a number of ways that Gombrich points out (1979): artists, like natural scientists, are often interested in that 'sense of order' that makes the sensible world accessible to our rationality; art has often been pursued as a science, in the relation between theory and a practice; and artists, like scientists, are often interested in 'truth' or the production of veritable reconstructions of the 'real' world.

Gombrich begins Art and Illusion with a discussion of Constable's conception of painting as 'natural philosophy'. Constable saw himself as a scientist, investigating the natural world by drawing and painting it. Drawing objects is a quite effective way of learning about those objects. If one performs such 'experiments', one cannot help but
discover certain features about the physical world. Gombrich draws attention to
constable’s success in his scientific pursuits:

Indeed, there are artists who think the field to which Constable devoted his
scientific endeavors has been fully investigated by now and that they must
turn to different areas for experiment. Instead of exploring the visible world,
they probe the mysteries of the unconscious mind or test our response to
abstract shapes. Compared with these hectic activities, Constable’s painting
of Wivenhoe Park looks so natural and obvious that we are inclined to
overlook its daring and its success. We accept it as simply a faithful record
of what the artist actually saw in front of him — “a mere transcript of nature,”
as paintings of this kind are sometimes described, an approximation at least
to that photographic accuracy against which modern artists have rebelled.
Let us admit there is something in this description. Constables’ painting is
surely much more like a photograph than the works of either a Cubist or a
medieval artist. (Gombrich 1960: 34)

Although there are many different ways in which works of art are representational, this is
certainly the conception of art as a transcript of nature that Plato, Descartes, Hume and so
many others had in mind when they compared the mind to pictures in its capacity to
represent the world. 14 In this way, a good picture is like a photograph. As Constable is
purported to have said, “The art pleases by reminding not by deceiving.”15

But even in representational art forms such as these (as opposed to photography
and film), the work of art is not representational because it resembles the world. Rather, it
is able to resemble the world because it is representational. The representational

14 Although the scope of this chapter does not allow a full discussion of representation in art, and I
have focused on the less ‘subjective’ forms of representational art, the theory of representation that I
present also allows for subjectivity. History-laden structures in media can represent subjective states
such as thoughts, imaginings, etc. just as they can represent the world. My painting represents my
ideas or imaginings when they are caused by my ideas or imaginings, in that I use the medium in
such a way as to ensure that structural features the structures in the medium co-vary with structural
features of my ideas or imaginings. I paint the square red instead of blue because I imagined a red
square not a blue square. Or perhaps I paint with these colors and not those, because I was feeling
happy. Had I not been feeling so happy, I would have used other colors. In addition, many history-
laden structures have multiple causes, some direct and some indirect. Thus a work of art can
represent, at one and the same time, objects in the world and the artist’s subjective states since both
may be causally responsible for the representational structure (the work of art) having the structure
that it has.

15 Gombrich (1960: 35) attributes quote to Constable.
structures in the medium are causally connected to the objects that they represent, albeit via the artist, and the medium (when used by a sufficiently skilled artist) can be sufficient to ensure that key features of the artistic structure will co-vary with key features of the object of art. Even a young child can represent her family veritably with her stick figures, provided that she intends that key features of her painting co-vary with key features of the people that she portrays (eye and hair color, for example), and that her abilities are sufficient for her to fulfill these intentions. Thus, if we are familiar with children’s art, when we look at a child’s picture of her mother, we can quite accurately infer that her mother has certain physical characteristics (for instance, long brown hair and green eyes).

Goodman makes a similar point in Languages of Art, focusing on the role of practice or custom in using media in certain ways:

Our addiction, in the face of overwhelming counter-evidence, to thinking of resemblance as the measure of realism is easily understood in these terms. Representational customs, which govern realism, also tend to generate resemblance. That a picture looks like nature often means only that it looks the way nature is usually painted. Again, what will deceive me into supposing that an object of a given kind is before me depends upon what I have noticed about such objects, and this in turn is affected by the way I am used to seeing them depicted. Resemblance and deceptiveness, far from being constant and independent sources and criteria of representational practice are in some degree products of it. (Goodman 1968: 39)

Resemblance should be considered a product of representation. But, because it is so often the result of representation, it is often important in interpretation. When we see two things as resembling one another, it is quite natural to interpret one as standing for or representing the other.

Yet, resemblance is only one principle of interpretation, for we often interpret one thing as standing for another even when there is no apparent resemblance. I have tried to present a case for a related, but more general, principle of interpretation: reasoning from
history-laden structures. Wax tablets, film, video, wet sand, plaster and snow, and all forms of non-literary representational art (including painting, drawing, sculpture, dance, music and theater) are media for representing the world. These media represent the world when the structures in the media are causally connected to objects in the world such that the media ensures that key structural features in the medium co-vary with structural features of those objects. Often, but not always, the medium is sensitive to objects in ways that will produce a relation of resemblance between the original object and the representation, when they are viewed or heard by ordinary people.

I am a camera.\textsuperscript{16}

Since ancient times, philosophers have used such media as analogies for the brain or mind, in its capacity to represent the world. This is an idea I will explore in some depth in the next chapter. I conclude this chapter by suggesting that these analogies have been so persistent in philosophy of mind because they are all representational history-laden structures just as the brain is, just as the mind is, or just as people are. These models all capture a very important truth about people: we are a very sensitive medium (like wax, photographic film, or paint and canvas when used in certain ways) in that when we have normal sensory experiences, the things that cause our experiences alter us in quite regular ways. The brain is a medium in which the world is represented. I will consider, in chapters three and four, specific issues that arise when the medium theory is used to understand neuro-representation; in the remaining pages of this chapter, I want to show, in broad outline, how the brain can be understood as a medium of representation.

\textsuperscript{16} This is taken from the title of John Van Druten's 1952 play based on Christopher Isherwood's \textit{Berlin Stories}.
Using the medium theory of representation, history-laden neuro-structures represent objects (or states of affairs) in the world in the medium of the brain if objects in the world were causally responsible for the history-laden neuro-structure and if the medium of the brain ensures that key structural features of the neuro-structure co-vary with key structural features of objects (or states of affairs) in the world. By considering key structural features of objects in the world, and the ways in which the brain ensures that structural features of brain states will co-vary with structural features of objects in the world, we can understand how it is that we represent the world. Alternatively, by considering structural features of brain states, and the ways in which the brain ensures that they will co-vary with structural features of objects in the world, we can interpret the brain states as representations of objects in the world.

People learn from experience and people remember things. This requires that people be sensitive to sensory perception. We are restructured in predictable ways by the things we come in contact with, and the ways that we are restructured are determined by the properties of the medium of the restructuring — our brains. People and other vertebrates undergo restructuring when they process sensory stimuli. A baby first begins to have perceptions of the world around her. As the visual stimuli are processed, her visual cortex and related parts of her brain are restructured accordingly. Gradually she begins to see things as we do — perhaps at first she sees boundaries, color splotches and light sources, then objects, faces, etc. Eventually such restructuring will allows her to respond differentially to things that she perceives: smiling at smiling faces, reaching for things she sees as objects, pulling at beards, etc. This suggests that there is some covariance between features of history-laden neuro-structures and features of the
observable world. The brain would appear to be a sensitive medium for representing the sensible world.

In addition, the structure of the brain constrains how the brain can be reshaped by events that occur to us, just as the properties of the medium in art constrain the ways in which structures in the medium could represent different objects. Although brains are reshaped by a person's sense perceptions, the prior structure of the brain determines the ways in which it can be restructured by those sense perceptions. Our capacities to learn different motor skills, learn languages and to remember things are due to our brain's inherent structure, prior to the events from which we learn. The prior structure of the brain allows it to be reshaped in certain ways, in response to different stimuli. If the structure is different, then the ways in which it can be restructured also differ. Thus, if a child has severe brain damage to a certain part of the brain, then the child may never be able to learn certain types of things. If there are small structural abnormalities in a certain part of the brain, this can constrain the way that it can be restructured. When certain brain structures are malformed, a person may not have the capacity to develop in certain ways or the capacity may be limited. In everyday life, our past experiences, and thus, past brain structures, affect the way that the events occurring to us will affect us. The ways in which our sensory perceptions reshape our brains are constrained by the prior and ever-changing structure of the brain.

Vertebrate brains are media that get restructured in response to sensory stimuli. Later, these new history-laden structures allow them to respond in new ways to like sensory stimuli. This is learning. Learning requires sense organs¹⁷, a capacity for behavior, and a medium of representation to connect the sense organs and the behavior.

¹⁷ Or functional analogues in things such as computers.
The medium must have representational capacity, since in order to produce new behaviors in like circumstances (learning) key features of the medium must co-vary with key features of the circumstances which caused the representational neuro-structures. The fact that we learn from our past experiences requires that brain structures represent the world in the kind of way that history-laden structures do.

As with other history-laden representational structures, if we know enough about the medium, then we can quite successfully interpret even a novel representational structure. Although our neurosciences has not yet developed (and quite likely never will develop) to the extent to which this is possible from brain structures alone, we have no trouble making such interpretations from the behavior for which the brain structures are causally responsible. We know something of how people function as a medium for representing the world that they have experienced, just as we know something of how film works as a medium for representing objects in front of the camera. Likewise, we know something of how people’s behavior depends on their representations of the world, just as we know something of how photographs depend on the exposed film from which they are made. And (provided that our knowledge of the medium is adequate) just as we can reason from features of a completely novel photograph to features of the objects that must have stood in front of the shutter, so we can reason from features of novel human behavior, to features of the person’s past experiences.

Exposed film (in normal conditions) is a representation of whatever stood in front of the lens when the shutter opened. When the film is developed, the negative is a representation of the same thing. Thus, a person who has some knowledge of film as a medium of representation can examine the negative and infer from it to what was standing in front of the lens when the shutter opened. When a print is made from the negative, the
resulting photograph is another representation of the same thing that the negative and exposed film was a representation of — that which stood in front of the lens.

Like the exposed (but undeveloped) film, the negatives and the photographs are representations of that which the camera was focused on. Unlike the exposed, undeveloped film, the photographs are interpretable by ordinary human observers, in the sense that most ordinary humans with a minimal understanding of how photographs function as a medium of representation can make accurate inferences from the photographs to what stood in front of the lens. The negatives are also interpretable by human observers, but a greater understanding of or experience with the photographic medium is required, especially with color negatives (as opposed to black and white negatives). Undeveloped, exposed film cannot be directly interpreted by human observers.

For this reason, one might object that the exposed film is not a true representation. Yet, using the theory of representation I have developed, it is a representation, since it meets the two central conditions: a suitable causal history and a medium sensitive to the objects represented, in that it ensures covariance between structural features of the representation and structural features of the cause of the representation. Evidence that the second condition is met (despite our inability to interpret the undeveloped, exposed film) is provided by the fact that negatives and prints are made directly and mechanically from the exposed film. For this reason, one can imagine the science of photography developing to such an extent that a human being equipped with special glasses might be able to 'see' the images in the undeveloped film.

Consider the type of machine that one finds in 'one-hour developing shops'. The machine does most of the work in developing the photos from the exposed film — people are needed only to position the undeveloped film, to remove the developed negatives,
hang them for drying, and then position the negatives properly so that prints can be made from them. The machine utilizes the exposed film in producing negatives, and it utilizes the negatives in producing the photographs. The machine can produce negatives from exposed film and produce photographs from the negatives without engaging in interpretation. Yet, when the machine is working properly, interpretable representations in one medium (the photograph) are produced from uninterpretable representations in another medium (the undeveloped, exposed film).

By analogy, neural representations (history-laden neuro-structures) are like the undeveloped film, in that they are (given the current state of science) uninterpretable representations that are utilized in producing an interpretable product—behavior. History-laden neuro-structures meet the first requirement for representation in that there is a suitable causal history. We come to have the neuro-structures we have as a result of our interactions with the sensible world. They also meet the second requirement, that the brain is a medium sensitive to the sensible world, and thus ensures covariance between structural features of the neural representations and structural features of the represented objects. Evidence that the second condition is met (despite our inability to interpret the neural representations) is provided by the fact that representational behavior can be produced from the history-laden neuro-structures. For instance, if I look at a novel animal, then I can later behave in ways that represent that animal, by describing it or drawing a picture of it. If the medium of my brain did not ensure covariance between structural features of my neural representation and structural features of the animal, then it is not likely that my brain structures could later be utilized in producing interpretable representations of that animal.
For this reason, one can imagine a world in which neuroscience has developed to the extent that a neuroscientist equipped with special machines might be able to interpret neuro-structures, in so far as she would be able to make accurate inferences about what sorts of things a person has experienced from certain structural features of the brain. Perhaps it is not unlikely that neuroscience will progress to the extent that neuroscientists will be able to make coarse-grained interpretations of brain structures, for instance, the interpretation of brain structures as characteristic of Francophones as opposed to an Anglophones, and thus, interpret such brain structures as representing French as opposed to English. It is unlikely (but perhaps possible) that neuroscience will progress far enough that more fine-grained interpretations of brain structures will be possible, for instance, the interpretation of brain structures as characteristic of someone who saw, at the age of seven, a tawny lion jump through a blue hoop in the center ring at the circus, and thus, interpret the brain structure as representing such an event.

But, whether or not neuro-structures can be interpreted, people utilize neural representations in the production of behavior. The very brain structures that have been shaped by a person's interactions with the sensible world are causes of people's behavior. This utilization of neural representation does not require interpretation any more than making prints from negatives requires interpretation of those negatives. But, as a result of the utilization, the person's behavior often reflects the person's past experiences. My answering the door when the doorbell rings reflects my past experiences with doors, doorbells and people. My sitting down and typing this chapter at the computer keyboard reflects my past experiences with language use, typewriters, computers, philosophy texts, graduate school, deadlines, etc. Our behavior is representational. Thus, we can interpret
it. We can infer from behavior to the kinds events that shaped the person such that she behaves as she does.

When we see a person get up and hop up and down while turning in circles at the end of a dart game, then we might take one of two lines of reasoning. 1) We can reason that the person must have had past experiences that explain why this is an appropriate behavior in this context; for instance, he was frustrated and he learnt to express frustration in non-destructive ways, or he was happy to have done so well, and learnt to express such happiness with victory dances, or perhaps, he wanted to draw someone’s attention, and he learnt that doing unusual things can work in this way. 2) Or we can reason that the person is crazy, on drugs, brain damaged, or ‘from a different planet’ — the person is in some important way a different sort of medium than what we expected. When we interpret behavior, as when we interpret footprints in the riverbed, we can err by making the wrong inferences from the structural features of the original object to the kind of object that it was, or we can err by misjudging the medium. But we can only misinterpret why people do what they do, because, in general human behavior is caused by our representations of past experiences in the medium of the human brain. Our behavior is caused by representational history-laden neuro-structures.

Any photographer, painter, holographer, playwright or poet must consider the properties of their respective media of representation, and how those media constrain the ways in which the artist can represent the world. Likewise, a neuroscientist must consider the properties of the brain as a medium — the neuroscientist must consider how our inherent brain structure constrains the way that it can be restructured, or reshaped, by sensory stimuli. The brain is a complex medium that imposes constraints on how the light that enters my eyes, and the stimulation of my ear drums and taste buds can restructure
my brain, as well as constraints on how we put all this disparate information together, and how we use it in our behavior. The past restructurings of my brain which constitute all that I have learned impose further constraints on what and how I will learn in the future.

When we consider structures in the brain apart from consideration of how the brain acts as a medium for representing the world, it can appear as if brain structures are not (on their own) representations of the world. Likewise, if we consider indentations in the riverbed apart from consideration of how fossilized mud acts as a medium for representing animals, then it will appear that the structures in the mud are not (on their own) representations of the animal that walked in the mud. Yet, there is an important way in which those structures point at or stand for the animal that caused them, simply because they were caused by that animal and none other, whether or not we interpret them that way. They represent the animals that caused them because the medium of representation ensured that structural features of the representation co-varied with structural features of the animals. Thus, whether or not we have adequate knowledge of the medium and how it can act as a medium of representation, the structures in that medium are representations, if only to a theoretically possible interpreter. The footprints do not suddenly become representations just when we come along and look at them, returning to their non-representational state when we turn our attention in another direction. They are representations so long as the medium ensures a covariance between structural features of the history-laden structures and structural features of the represented object to which it is causally and historically connected.

Sometimes we utilize history-laden structures, and at other times we interpret them, depending on our interests. When we interpret them, we interpret them correctly or incorrectly as representations of what which caused them; this depends on the context, on
our knowledge of the medium and on our ability to make correct inferences. But whether we can know whether or not we are interpreting them correctly or incorrectly as representations of what caused them is an entirely different problem. Further, our ability to utilize these representations as we do, in producing interpretable behavior which incorporates what we have learnt from past experience, depends on the fact that they actually are representations, whether or not we are able to interpret them, and whether or not we recognize them as being representations.

The analogy between minds or brains and representational art is a powerful one that depends on the simple and obvious notion that representations in a medium represent those things which produced history-laden structures in that medium. Some media are better than others for representing certain things. The human brain appears to be an excellent medium for representing the sensible world. Our capacity to engage in meaningful behavior, like our capacity to interpret others' behavior as meaningful, ultimately depends on the existence of history-laden structures, and thus, on media sensitive to different kinds of objects. All interpretable human behavior, from a child’s first word, to acts of murder and the composition of symphonies, requires the utilization of representational, history-laden structures. And no matter how we talk about people — as brains, minds, bodies or just as people — the more we learn about them as media for representing the world of their experience, the better we will be able to interpret them.
Chapter 3

USING THE MEDIUM THEORY
TO ACCOUNT FOR NEURAL REPRESENTATIONS

By thinking about representation as a history-laden structure in a medium, we can conceptualize the human brain as a medium adequate to ensure covariance (but not necessarily similarity) between structural features of history-laden neuro-structures (representations in the brain) and structural features of the causes of those neuro-structures (objects in the world). A neural representation (i.e. brain structure) is caused by activity in the brain such as occurs when a vertebrate interacts with the sensible world. Neural representations are utilized in the production of behavior, which is, as a result, interpretable as meaningful. Hence, people represent the sensible world in much the same way that a work of art, impressions in the snow, or undeveloped film represents its subject.

Yet, despite its strengths, there are obstacles to the analogy between neural and artistic representations. In this chapter I consider three objections to the medium theory of representation as it applies to the brain: 1) the brain is neither a single-use nor single-exposure medium; 2) the problem of misrepresentation; and 3) the problem of determining the object of representation (why neural states represent external objects rather than states of our sense organs). Two other problems, the disjunction problem and Twin Earth considerations, will be addressed in chapter 4 in the context of a discussion of linguistic representations.

Single-Use, Single-Exposure Media

The most obvious problems with the analogy between neural representation and natural or artistic forms of representation are that most media of representation are limited
to a single use and they require a single exposure to the object of representation. When you paint a new picture over an old one, press a new seal into the wax, or record a new event on video, you lose the first picture or impression. In addition, the impression is made during a single episode or event. The wax is impressed at a certain time and place, the camera works instantly. The examples that I have used in developing the analogy (signet rings, footprints, photography, etc.) tend to involve only simple relations between media and objects: the history-laden structure is produced through a single event in which the medium responds to the object, and in a medium that can accommodate only a single representation at a time. Brain states do not seem to function like that — countless different things are represented in the brain, and often it seems that we require many experiences with objects or type of events before we can utilize neural representations of them.

Of course, there are some non-neural representational structures that result from multiple events, each of which involve the medium's exposure to an object, or to different objects. Consider time-lapse photography, or other photographic effects produced through multiple exposures. A painter may require the model for a number of sittings. And by extension, landforms shaped by centuries of exposure to the wind and rain may represent years of interaction with the elements, and by studying the current structure of the landform experts can make inferences about the area's history. However, these do not provide a good model for how the brain acts as a multiple-use, multiple-exposure medium of representation.
Such disanalogies have long been recognized and addressed by those who have built on the analogy between representational art and mind. The Stoics compared the representational capacity of the soul to impressions in wax:

So, according to them [the Stoics], a presentation is an impression in the soul. And they differed immediately about this. For Cleanthes took 'impression' in terms of depression and elevation — just like the impression on wax made by seal-rings. But Chrysippus thought that such a view was absurd. For first, he says, this will require that when our intellect has presentations at one time of a triangle and a tetragon, the same body will have to have in itself at the same time different shapes — triangular and tetragonal together, or even round; which is absurd. Next, since many presentations exist in us at the same time the soul will also have many configurations. This is worse than the first problem. [Chrysippus] himself speculated, therefore, that 'impression' was used by Zeno to mean 'alteration'; so that the definition becomes like this: "presentation is an alteration of the soul"; for it is no longer absurd that the same body at one and the same time (when many presentations exist in us) should receive many alterations. For just as air, when many people speak at once, receiving at one time an indefinite number of different blows, also has many alterations, so too the leading part of the soul will experience something similar when it receives varied presentations. (Sextus Empiricus, Adversus Mathematicos: 7.2227-236)

Their solution to the problem involves understanding representations as 'alterations' rather than 'impressions'. This suggests a kinship with the medium theory of representation in its emphasis on systematic change as opposed to resemblance. If representations are considered to be 'alterations', then it is much easier to find analogies between the human mind and other media which can undergo many different alterations simultaneously without interference.

Descartes, who is often held responsible for more recent analogies between the mind and forms of representational art, also recognized the limits of the analogy with respect to the brain. In The Passions of the Soul, Descartes presents a conception of brain states which avoids the problem of how neural representations can be caused by

---

1 I am grateful to Brad Inwood for drawing my attention to this passage.
2 Translated as B 8 in Inwood and Gerson (1988); see also 7.372-373 (B 9).
sensory perceptions without being replaced or obscured by more recent impressions. Indeed, his account shares many features with contemporary connectionist accounts of the brain. Rather than connections between neurons, however, Descartes concerned himself with patterns in the pores or cavities of the brain through which the animal spirit (a corporeal substance) flowed. Whenever we look at something the particular way in which the eyes are stimulated pushes the animal spirits in different patterns through the pores of the brain, and they create little paths. The spirits flowing through the pores push on the pineal gland (the seat of the soul), and likewise, when the pineal gland moves, it pushes the spirits back through these pores. When the animal spirit flows back in the same patterns that were caused by certain sense perceptions, it results in memories, dreams or ideas, which bear resemblance to the original sensory images that caused those patterns in the brain. Thus, the brain has an unlimited number of states, many caused by sense perceptions, and each corresponding to a position of the pineal gland, which corresponds, in turn, to a mental, representational state:

42. How we find in our memory the things we want to remember. Thus, when the soul wants to remember something, this volition makes the gland lean first to one side and then to another, thus driving the spirits towards different regions of the brain until they come upon the ones containing traces left by the objects we want to remember. These traces consist simply in the fact that the pores of the brain through which the spirits previously made their way owing to the presence of this object have thereby become more apt than the others to be opened in the same way when the spirits again flow towards them. And so the spirits enter into these pores more easily when they come upon them, thereby producing in the gland that special movement which represents the same object to the soul, and makes it recognize the object as the one it wanted to remember. (Descartes, Passions: AT XI: 380)

Descartes' conception of neural representation shares two important features with contemporary accounts: distribution of the representation throughout the brain, and

---

3 For a discussion of the one-to-one correspondence between mental states, positions of the pineal gland, patterns of animal spirits in the brain, see articles 34 and 35 in Passions.
superposition of neural representations (since the same channels and pores would be
involved in many representations). And like Chrysippus', Descartes' account of neural
representation involves systematic changes to the brain structure, changes which do not
necessarily produce resemblance between the object of sensation and the 'impression'.

Descartes is explicit in his criticisms of resemblance theories of representation in
Optics, discourse four, which is on the topic of the senses 'in general':

We must take care not to assume — as our philosophers commonly do —
that in order to have sensory perceptions the soul must contemplate certain
images transmitted by objects to the brain; or at any rate we must conceive
the nature of these images in an entirely different manner from that of the
philosophers. For since their conception of the images is confined to the
requirement that they should resemble the objects they represent, the
philosophers cannot possibly show us how the images can be formed by the
objects, or how they can be received by the external sense organs and
transmitted by the nerves to the brain. (Descartes, Optics: AT VI: 112)

Descartes then gives a number of reasons why full resemblance is not produced, both in
artistic representations such as engravings and also in the brain. It should be noted,
though, that Descartes did think some resemblance would be produced, at least with
respect to visual representation⁴. He argues, however, that perfection in representation
does not depend on complete resemblance between the image and the represented
object. "Now we must think of the images formed in our brain in just the same way, and
note that the problem is to know simply how they can enable the soul to have sensory

---

⁴ In Descartes' descriptions of representation in the brain, although he often stresses that
resemblance is not necessary, he also occasionally used analogies such as the traditional wax and
seal metaphor, which do produce resemblance, see for instance, Descartes' discussion in Rules: AT
X: 414-416. See also the discussion of visual perception in Man: AT: XI 175. In this mechanistic
description of vision, some resemblance is produced, although resemblance is not produced for all
aspects of vision (for instance, color, distance to the object) nor for any other sense, such as sound,
touch, taste, pain, emotions, etc. When these passages are taken in context, it becomes clear that
some resemblance may be produced but that it is a consequence of some particular mechanistic
accounts that are primarily intended to produce systematic covariance.

⁵ For more discussion of the failure of resemblance in neural representations, see Optics: AT VI:
131; Rules: AT X: 415-416.
perceptions of all the various qualities of the objects to which they correspond — not to know how they resemble these objects." (Descartes, Optics: AT VI: 113)

Descartes saw that representations in the brain were necessary to account for animal behavior (which required perception, memory and learning), if one wanted to avoid appealing to notions such as conscious states, awareness and reason (which were denied to animals on Descartes' account). These neural images do not resemble but systematically correspond to the various qualities of their causes and, in addition, correspond systematically with their effects (for instance, movements of the body are caused by images that do not resemble but correspond to movements in the brain). On Descartes' account, animals and humans share two important corporeal structures in the brain: a 'common' sense which receives sensory information; and the corporeal imagination which actively uses received sensory information, as well as images formed in the past, in the production of movements appropriate to the situation. Thus movement is produced mechanistically from sensory input in humans almost exactly as it is produced in animals. The difference between people and animals is that we also have an incorporeal soul which is moved by the 'common' sense and moves the corporeal imagination, through the pineal gland. Thus, in humans, but not in animals, the brain activity involved in sense perception and movement produces conscious states via the pineal gland.

Like Descartes' theory, connectionist models of the brain make central the ability to use sensory stimulation as input, learn from it, and produce different behaviors as output. This means that with respect to learning, memory and the acquisition of knowledge, Descartes (at least in his more mechanistic moments) concerns himself with the question 'How is sensory information utilized in the production of systematic behavior?' instead of

---

asking traditional philosophical questions concerning cognition, understanding, meaning, truth, etc. This is, of course, also the question that motivates much neuroscientific research. Indeed, after giving a detailed account of brain functions on a connectionist model that focuses primarily on sensation and motor control, Patricia Churchland ponders:

"Why should philosophers care about a theory whose general shape is dictated by problems of sensorimotor control?"

"This is a fair question, and its answer depends on thinking about the human mind-brain as something that evolved from simpler nervous systems, then increasing sophistication in nervous systems is, in this broad sense, understandable as an increasingly sensitive means of controlling behavior on the basis of sensory information. (Churchland 1986:450)

Any account of learning, that is, the alteration of behavior in response to past experiences, must suppose that a vertebrate represents key aspects of the world which he senses. When a vertebrate is conscious and interacting with the world, this interaction causes alterations to the structures of the brain. These systematic alterations can be understood as neuro-representations which affect future patterns of neural activity in response to sensory input, and so alter the vertebrate’s behavior in systematic ways. Representations are utilized in neural activities (including computation, movement, remembering, talking, imagining, etc.) These neural activities consist of the activation of neural representations. Thus, any neuroscientific account of human behavior is ultimately an account of how the brain acts as an active medium of representation, sensitive to perception and capable of using those representational structures in the production of behavior.

On a connectionist model of the brain, representations are either patterns of neural activity (active representations, such as when we are perceiving) or ‘connection weights’ which determine the patterns of neural activity given sufficient stimulation (in stored or passive representations). Robert Cummins summarizes the differences between what he
calls orthodox representationalism and connectionism. He describes the connectionist representation as "activation patterns" or "connection weights"; connectionist performance as "computation of representations via spreading activation as determined by connection weights", and learning as "systematic changes in the connection weights" (Cummins 1991: 97-98). Similar descriptions of neural representations are given by Patricia Churchland (1986), Goschke and Koppelberg (1991) and Haugeland (1991).

Neural representations, then, can be thought of as the activation patterns in a neural network caused by such brain activity as occurs when a person is conscious; alternatively, as the patterns of connection weightings which become activated given sufficient stimulation. It is difficult, if not impossible, to demarcate such representational structures; however, their presence is evident when they are utilized in the production of interpretable behavior. The difficulty in individuating these structures is due to several facts: that they are distributed throughout the brain; that they are superposed on each other; and that they are constantly changing as a result of sensory interaction with the world and their own utilization.

For instance, suppose that a complex pattern could be picked out as a neural representation of cats. By hypothesis, the pattern of connections was caused by the neural activity produced on the occasions that a person saw cats. Different patterns of neural activity would have been produced on each occasion; however, where there was overlap, relatively more stable neural patterns would have begun to develop. The slightly more stable patterns would have resulted in even more overlap on future occasions, since such stable patterns are more easily activated as whole systems. In this way, the neural representations can result from multiple exposures. In most cases, the more that a person is exposed to cats, the stronger the neural representation would be.
This understanding of neural representation suggests that stable representations from single exposures would be possible, but relatively rare. The more unusual an event, the more likely it is a recognizable neural pattern would result from a single exposure, since it would likely create a greater difference in the connections and thus a new pattern might stand out against the old. The longer the period of time from the point of representation to the point of utilization, the more likely it is for the pattern to be eroded by other, overlapping brain activities. This degradation would be reduced if the neural representation were to be frequently utilized, since the activity would reinforce those patterns, or at least the portions of the patterns that were utilized. In addition, two or more patterns that are frequently activated at the same time would become associated, since the sum of the patterns could become a pattern itself. A consequence of this is that when one of those patterns is activated, the pattern of activity could result in some activation of one of the associated patterns.

Not only does the media model of representation provide an understanding of neural representations that is compatible with connectionist models of the mind, but neural representations, on this model, would have many characteristics that we associate with ideas and memories. Many stable ideas or concepts develop only after repeated exposure to an object. Less often, a concept or idea becomes firmly rooted even after a single exposure. The more unusual the object or concept is, the more likely that this will occur. With time, some concepts or memories get lost, details get lost, or they become vague. Memories of more unusual events may be more stable. The more that happens to us, the more activity that we engage in, the more likely we are to forget. We can minimize the loss by utilizing the ideas or memories, replaying them, practicing, etc. Still, the aspects or features that we don't practice or frequently remember or draw on can be lost, such that only the most important details remain firm. In addition, if two ideas or events frequently
occur with one another, then they can become associated, such that the seeing or thought of one can trigger the thought of the other. If they cease occurring with one another, then the association will gradually fade or be replaced by new associations.

Connectionist models of the brain stress the close relationships between different types of learning, different types of memory, different types of habit formation, and different types of cognition. The terms with which we talk about different learning, memory or other cognitive processes need not correspond to neural processes. Patricia Churchland comments:

Learning and memory are at the dead center of cognition, if anything is, and as their categories are revised and redrawn, the theoretical landscape of higher functions is undergoing tremendous transformations. The general category of learning has already fragmented into a variety of kinds of process, and indeed the term "learning" is now often replaced by the broader and less theoretically burdened expression "plasticity".

Among kinds of nervous system plasticity already believed to be distinct phenomena are habituation, sensitization, classical conditioning, operant conditioning, imprinting, habit formation, post-tetanic potentiation, imitation, song learning (in birds), one-shot learning to avoid nausea-producing foods, and cognitive mapping, in addition to which are the apparently high-level phenomena distinguished in terms of what is learned, such as learning language, learning who is a conspecific, learning to read, learning social skills, learning mathematical skills, learning to learn more efficiently, learning to lower blood pressure, and heaven knows what else. Whether these involve common or distinct mechanisms, and how many distinct processes there really are, remains to be discovered. (Churchland 1986:151-152)

What remains certain is that these many varied processes, whether or not they turn out to be distinct neuro-physiological processes, must utilize representational structures in the brain. The term, 'plasticity', brings to the forefront an important feature of the brain, but also of representation itself. The brain, like any medium of representation, undergoes alterations as a result of interaction with objects in the world. It is 'plastic', that is, its structure can be altered.
The brain is a very good medium for representing the world of our experience. It shares with other media of representation the capacity to produce systematic alterations which co-vary with the represented object — systematic alterations that need not produce resemblance. It is unlike many other media of representation in that it is a multiple-exposure, multiple-use medium. Thus, many neural representations will result from multiple exposures to objects or events, although more unusual events may result from single exposures. The reliance of the medium theory of representation on alteration rather than resemblance or straightforward signification gives it an advantage in accounting for the multiple-use, multiple-exposure representations, especially given a connectionist model of the brain. The conception of representation as alteration is neither novel nor particularly unusual, and has been used in the past to explain mental or neural representation, often resulting in models of the brain or mind that are interestingly similar to connectionist accounts. Since the brain is the only known intermediary between the past events of which our overt and mental behaviors are representations, the simplest explanation is that the brain is, in fact, a medium of representation.

Misrepresentation

The theory of representation I have presented is a causal covariance theory: things represent that which they are causally connected to and with which they co-vary. In this way it might seem to share with other causal theories of representation a difficulty in accounting for misrepresentation; consider Locke's covariance theory, the causal-informational theories offered by Dretske (1981, 1986) and Fodor (1990), and also Millikan's teleological theory of intentional devices including representations (1984). The

---

7 Even in the brain, some resemblance may be produced as a result of systematic alteration. Indeed, Michael Kosslyn (1983, 1994) presents evidence in favor of some resemblance resulting from neural representation in visual perception and mental imagery.
problem with covariance theories or causal theories of representation is that it is precisely
covariance or a suitable causal history that makes the structure representational. Thus, in
so far as a representational structure fails to co-vary with that which it is supposed to
represent, or fails to have a suitable causal history, then it is not representational at all.
The problem that there is no room for misrepresentation in causal or covariance theories of
representation (as well as possible solutions) has been much discussed, see for instance,

Before discussing possibilities for misrepresentation in the medium theory, it is
important to distinguish misrepresentation from misinterpretation. As discussed in the
previous chapter, in simple cases of single exposure representation, the represented object
is the object that is causally responsible for the representational structure (the brass signet
ring that actually made the impression in the wax). The interpreted object is the object that
an observer interprets the structure as representing (Johannes’ almost identical gold signet
ring which Tanya believed to be responsible for the impression in the wax). The accuracy
of interpretation depends on the medium’s sensitivity to the types of objects that caused
the representational structure, and on the interpreter’s familiarity with the medium, her
knowledge of the context and her ability to reason from a representational structure to its
causes.

Clearly, the interpreted object and the represented object need not be the same —
people make mistakes when interpreting. Misinterpretation occurs when a person makes
incorrect inferences from a representational structure to the represented object, thus
interpreting it as a representation of something other than that which caused the structure.
Misinterpretation can result from insufficient knowledge or from faulty reasoning. For
instance, if I see footprints in a dried riverbed, then I might interpret them as caused by,
and thus representing, a prehistoric animal. However, it may turn out that anyone with better knowledge of the medium of mud and the changes it undergoes through time would have immediately seen that the footprints could not have been caused more than a thousand years ago. In this case, my misinterpretation is due to a lack of familiarity with the medium of representation. If, however, I look at a photograph of Cole, I might reasonably but incorrectly interpret the photograph as a representation of his twin brother Colin. Although my knowledge of the medium of photography is sufficient to infer that the photo is of someone who looks like Colin, I am lacking other information relevant to who the photo represents — I do not know who stood in front of the camera. Faulty inferences may also result in misinterpretation. I might accurately infer that the photo represents someone who looks like Cole or Colin, but fallaciously reason that it could represent only Cole or Colin. I may reason that since the date on the back of the photo is a time at which Cole was dead, it must represent Colin, failing to consider the possibility that they were not twins but triplets. Thus, I misinterpret the photo as a representation of Colin, when in fact, it is a representation of Leo (the third brother or even an unrelated lookalike.)

Although we are still unable to interpret neuro-representations, in theory, neuro-representations could be similarly misinterpreted. If a future neuro-scientist’s knowledge of the medium of the brain is sufficient for her to interpret neuro-representations, then she might also misinterpret neuro-representations in the same ways. For instance, although her familiarity with the medium is often sufficient for her to accurately interpret the causes of neuro-representations, she may lack knowledge with respect to the effect of time on some neuro-structures. Thus, she may misinterpret a neuro-representation as representing a person’s childhood friend, when in fact it represents the person’s childhood imaginary friend. If such a representation had been caused more recently, she would have
been able to distinguish it from the representation of a flesh and blood friend, but the degradation caused by the passage of time made the imaginary-friend-representation look like a degraded flesh-and-blood-friend representation. Alternatively, the neuro-scientist might correctly interpret a neuro-structure as representing water, yet, in her reasoning, she may fail to consider the possibility that it was caused by Twin Earth water (XYZ) rather than H₂O. Thus, she misinterprets it as a representation of H₂O, when in fact, it represents XYZ.

A future neuro-scientist who can interpret neuro-representations is almost as fantastic as Twin Earth itself. We cannot now interpret representational neuro-structures, and it is unlikely that we will ever be able to do so, at least not in such a fine-grained way. But we can and do interpret behaviors caused by representational neuro-structures — behaviors which are themselves representational, and thus subject to interpretation and misinterpretation. Consider an ordinary behavior in Toronto: a person walking along the street, reaches a corner where there is a crosswalk, points (across the intersection) and walks across the street. From this behavior, we might reasonably infer that the person is utilizing representations, for instance, of the crosswalk, of the road, etc. We might reasonably assume that these representations are caused by his past experience with crosswalks like this one. Indeed, if you interpreted my behavior when I perform such an act as representing my past experience with crosswalks such as these, you would interpret my behavior correctly. I have had past experience with crosswalks, which caused a neuro-representation of crosswalks which is then utilized when I recognize a crosswalk and then behave in this way. (Indeed, the fact that I do recognize crosswalks and behave in this way in such circumstances is the best evidence we have that I have such a neuro-representation.) However, we might misinterpret another person's behavior when he, inexperienced with Toronto crosswalks, happens to behave in much the same way.
Suppose he is not utilizing representations of crosswalks at all. Rather, when he reaches the corner, he happens to point across the street to his destination (or an interesting building) in order to point it out to his companion. Cars stop when their drivers misinterpret his pointing. Then, he and his companion cross the street, marveling at the consideration of Toronto drivers.

We misinterpret people's behavior for all sorts of reasons. Not only because similar behavior can be produced in different ways from different representational neurostructures, but also due to inadequate knowledge of the medium, a lack of other relevant information, and through faulty reasoning. As I discussed in chapter two, misinterpretation most often occurs when we under- or overestimate the medium's capacity for representing, or when we make unwarranted inferences.

Unlike cases of misinterpretation in which the fault lies with the interpreter, or in the act of interpretation, cases of misrepresentation involve a fault or problem in the representational structure itself, prior to any act of interpretation. This suggests that the medium theory, like any other covariance theory, must have a difficulty accounting for misrepresentation. Representational structures in an adequate medium of representation must successfully represent that which caused the alterations to that medium whether or not they can be correctly interpreted. And indeed, it is likely that that complete misrepresentation rarely, if ever, occurs naturally. Two types of misrepresentation can, however, be accounted for — misrepresentation through human agency or deception and partial misrepresentation.

The clearest examples of misrepresentation are ones which involve human agency, and more specifically, human deception. Arthur Conan Doyle, for instance, wrote a book on the subject of fairies which included a number of photographic plates of fairies (1922).
These plates misrepresented fairies as existing. Doyle used intentionally doctored photographs such that they would appear to be photos of fairies. He could practice this deception because the medium of photography is generally adequate to ensure covariance between features of the object of representation and features of the photographic image, and thus, is generally adequate to ensure that images of fairies do not naturally appear in photographs. It is only when a deceptive agent interferes, intentionally altering the photographic process such that the end result will appear to be an ordinary photograph even though it is not. Doyle was aware that people knew enough about photography to be able to interpret photographs as good evidence of what exists in the world. With this knowledge, as well as his photographic expertise, he could produce photographs in which fairies were misrepresented as existing. Indeed, Doyle used his knowledge of photographic processes as well as procedures of authentication to argue that the photos were not doctored.

Fakes, counterfeits and forgeries often involve this kind of misrepresentation. Although they can in theory be produced through human error, by far the best examples of misrepresentation are produced through human deception. But it is only because the medium used for the misrepresentation is, in general and in similar contexts, a successful medium for representation that it can be used successfully to misrepresent. Thus, a watercolor by Sulamith Wülfing cannot misrepresent an elf as existing, since there is no expectation of covariance between features of her fantastic paintings and features of the world. A faked seal on an historic document, on the other hand, can misrepresent, since we have expectations that the wax impression would be causally connected to the original signet ring. Wax is, after all, a fairly good medium for representing signet rings. In misrepresentation, as in representation, it is the character of the medium that determines how successful the representational structure will be at representing or at misrepresenting.
For the most part ordinary misrepresentation involves human agency, deception and error; however they do not seem to be the types of misrepresentation that Dretske (1986) and Cummins (1989) are concerned with. Rather, the difficulty for covariance theories is with a passive or agentless misrepresentation. Many of the difficulties that other causal or covariance theories of representation encounter are overcome when representations are distinguished from signs and symbols. In the medium theory of representation, representations differ from signs and symbols in that they involve both a medium and covariance between structural features of the represented object and the representational structure, rather than direct and simple covariance between a structure and that which it stands for.

These features give the medium theory an advantage over other covariance theories of representation in explaining misrepresentation, since it allows for partial error, rather than all-or-nothing representation. A history-laden structure in a medium can be understood as partially misrepresenting an object, or misrepresenting a feature of the object, if the medium generally ensures covariance between structural features of the representation and structural features of the object of representation, but due to unusual features of the situation, things go awry (but not so far awry that it is a complete failure as a representation). For instance, photography is, for the most part, a sensitive medium for visible objects, but things can go wrong in the mechanical process on a certain occasion, for instance, if my film gets exposed to a tiny ray of light through a crack in the camera case, the photograph may misrepresent its object as having a white splotch in the middle of it; in other cases, the photograph may misrepresent a person as having red eyes.

A consequence of the medium theory of representation is that the more experience one has with a medium or of the object of representation, the less likely one is to be fooled.
or taken in by a misrepresentation. The photograph may misrepresent a person as having red eyes, but most people correct for this misrepresentation when observing the photograph. After all, although color photography is in general a good medium for representing people's eye color, it is not at all unusual for flashes to reflect off the retina causing red eyes to appear in the resultant photograph. Other partial misrepresentations may go unnoticed. Suppose that there is a slight flaw in an audio tape due to a surge in power during recording. The tape misrepresents a short burst of notes as slightly higher and faster. It may be that only an expert in sound recording or an expert in symphonic music can recognize the flaw when listening to the tape. The ordinary person may be completely oblivious to the partial misrepresentation.

On this theory, what counts as a partial misrepresentation may seem arbitrary. After all, the characteristics of the medium of representation determine which features of the representational structure will co-vary with which features of the represented object. With respect to color, the photographic medium is imperfect, especially when used by inexperienced photographers. Thus, it may seem that the red eyes and the flaw in the audio tape are more like cases of misinterpretation in which the interpreter misjudges the medium's capacity for representing an object. Yet, in such cases of misrepresentation, the medium is not generally insensitive to such features. For instance, photographs are generally sensitive to color and one can compare the specific 'red eye' discoloration to more general discoloration due to over- or under-exposure. Photographs rarely capture the exact colors of their subjects, and it would be a mistake to interpret the colors in a photograph as exactly true to life. Yet, only the eyes are shown as bright red when they should be blue or brown. The small flaw stands out against the success of the rest of the representational structure. In all other respects the photo may be an excellent representation of a person.
It is possible for things to go awry with human representation as well, allowing for natural or passive misrepresentation. Although much of the working of human memory remains a mystery, for the most part, the human brain is quite a successful medium for representing the world of our experience. Thus, a person can often describe an event that they have witnessed, or a place that they have been. Yet, in principle, things can go wrong in the brain during the occurrence of an event, such that although for the most part, the event is accurately represented and can later be described, some details are incorrect — a neurological equivalent to red eyes. That things can go awry in the representational stage (rather than in the later stage of recalling the event) is suggested by the fact that a person’s mood or condition (distracted, drunk, sober, exhausted, etc.) at the time of an event can affect how well they remember it. If this type of neurological malfunction can occur, then just as a photograph can accurately represent a person except for the red eyes, neurological structures may accurately represent things or events in most but not all respects.

The most interesting cases of neural misrepresentation are cases of deceptive rather than partial misrepresentation, comparable to doctored photographs or forgeries, and thus involve human agency. Deceptive misrepresentation requires that an agent who is aware that a certain medium is generally a medium adequate to ensure covariance intentionally interferes with the representational process such that a resulting representational structure appears to be a normal representational structure, but has been altered such that it appears to co-vary when it does not. Self-deception can be of this nature. A man, heartbroken when his partner leaves him, may intentionally create misrepresentations. He may think about past conversations or enjoyable activities shared with her, and focus only on the things that went wrong, even though at the time they were
trivial. Or when he sees her, he may try to convince himself how unpleasant their encounters are, even though she is being quite civil. He may look at beautiful pictures of her, or think about her, and tell himself that she is not really beautiful, that she is too skinny, too tall, not sexually attractive, etc. If he is successful in these endeavors, he may create in himself a representational structure which misrepresents his ex-partner and which functions, in determining his behaviors, as a normal representational neural structure (whether or not he also remembers that he actively deceived or convinced himself of her 'flaws'). Such active deception in creating misrepresentations can be an important part of 'psyching oneself up', settling down when excited, convincing oneself of something, developing a better attitude, etc. Closely related to these are less intentional forms of misrepresentation in which a person's mood or state of mind which they are only partially in control of causes misrepresentation, for instance, when one is blinded by love or hatred.9

Determining the Represented Object (External object or Internal State?)

One problem with causal, covariance theories of representation is that the complexity of the vertebrate perceptual system provides many sources of covariance. Not only do neuro-representations co-vary with the objects that we perceive, but they also co-vary with events in our sense organs.9 On what grounds can we say that the represented

8 Brain washing, suggestion and hypnosis can also be understood in the same way, but with an external agent. Once again, the clearest cases are those in which the agent is intentionally causing 'misrepresentations' by manipulating a representational medium (mind or brain) which he/she understands. Yet, deception can also occur more passively through error or oversight, through intentional behaviors but which are not intended to deceive or create misrepresentations.

9 Our neural representations also vary with the qualia of our perception. Yet, qualia are in no way represented by the neuro-representations, since the causal relation (if indeed there is one) goes in the wrong direction. A representational structure in a medium represents its causes. Qualia are not causes of representational brain structures. If anything, qualia are caused by neuro-representations, and thus, represent the same objects as do those neuro-representations, just as the print represents the same object as the negative or the exposed, but undeveloped film.
object is the external object that we perceive rather than the state of our sense organs when we perceive it?

So far, most of the neural representations that I have discussed are relatively simple in that they involve but one sense — visual cat representations, sound representations, etc. Indeed, features of these neural representations co-vary as much with the state of sense organs and with the sensible qualities of objects (coloration, shape, etc.) as they co-vary with the objects of perception. Thus it may seem that they are representations of the states of our sense organs or of sensible qualities of objects rather than of the objects of perception.

However, this problem is merely a result of considering perceptual representations in their simplest forms, focusing on one sense organ at a time. The problem resolves itself when we consider how do we come to distinguish between an object of perception (the cat that I see) and perceptual qualities (the shape and coloration of the cat); or consider how we merge these different perceptions (visual, auditory, tactile, etc.) into a multi-sensory representation.

It is difficult to imagine how we could make a distinction between a perceptual quality and the object of perception if we had only one sense. Imagine that the only sense you had was vision (or if you prefer, the sense of touch or hearing). If all your information about the world and yourself came through that one sense, on what basis could you distinguish between the objects of your perception and the perceptual qualities of those objects? What in your experience could generate a distinction between the visual object and the object of vision? Nothing would happen to you except changing visual fields which you would have little control over since you would have no sense of movement or touch.
Whether or not it is somehow possible to generate a distinction between perceptual qualities and object of perception with only a single sense, all vertebrates in fact have many senses: the five traditional senses (vision, hearing, touch, taste, smell), forms of internal sense (for instance, pain, thirst, hunger, and awareness of other internal conditions such as muscle fatigue and full bladder, etc.); some vertebrates even have senses that humans don't share, such as sonar. Whenever there are at least two different capacities for sensation it becomes in principle possible to distinguish between the object of perception and perceptual qualities.

Consider a perceiver with two senses: vision and touch. The perceiver experiences the world visually and by touch. Thus, given a capacity for forming neural patterns of activation, the perceiver can develop visual neural representations and tactile neural representations. If these happen to co-vary with one another (as visions of plush toys co-vary with feelings of softness) then a composite representation can result. Neither the visual representation nor the tactile representation is identical to the composite representation. If the visual representation is a representation of visual qualities, and the tactile representation is a representation of tactile qualities, what are the composite representations of? What is at the intersection between visual and tactile representations? The object that we see and touch. The composite representations are representations of objects that can be perceived by more than one sense, for instance, of objects that can be both touched and seen.¹⁰

---

¹⁰ When we cannot touch that which we see (and we see it as close enough to touch) then we may say it is an illusion or a hallucination. It is precisely when we only have one way of sensing something that we have difficulty distinguishing between the perceptual object and the object of perception. That is why color is so fascinating to philosophers: all you can do with color is see it. You cannot touch it, smell it, taste it, or hear it (synaesthesia aside). Vision is the preferred sense in philosophical discussion, and color is one of the few things that is only visual.

Two senses in two people who can communicate can produce the same distinction. If I can only hear a sound, and can't feel the rumbling or see the cause, or deduce the cause from other information, then I might think it is just a ringing in my ears. But, if someone else hears it as well,
Metaphorically, as we interact with the world, our multiple senses act as a prism acts on white light, breaking it up into many colors. But, then in processing the sensory stimuli at one and the same time, our brains act like a second prism, reuniting the colors back into white light. White light is composed of many colors, but it is not a bundle of colors. Rather, it is something that breaks into colors when we put a prism in its path. And, with a second prism, we can reconstruct white light. When we interact with the world, through our multiple sense organs, we break the world into visual, auditory and tactile components. We create perceptual representations that represent perceptual qualities. But then, we combine these perceptual representations back together. In doing so, we create composite representations that are not a bundle of perceptions, but a reconstruction of that reality which was broken into perceptual components by our sense organs. These composite representations do not represent perceptual qualities, but rather, as reconstructions of the object to which the varied perceptual qualities belong, they represent that object. We can distinguish between sensory qualities (sights, sounds, tastes and smells) and the objects that we take to be the cause of these sensory objects: the things we touch, see, hear, smell and taste.

then we are confident that there is an object of perception (the thing that rings) and not just a perceptual object (the ringing in my ears).
The following diagram helps to illustrate how the metaphor of the prism makes the point:

![Diagram of light passing through prisms](image)

<table>
<thead>
<tr>
<th>White Light</th>
<th>Prism 1</th>
<th>Colored Light</th>
<th>Prism 2</th>
<th>Reconstructed White Light</th>
</tr>
</thead>
<tbody>
<tr>
<td>The World</td>
<td>Sense Organs (Early stages of neural processing)</td>
<td>Representations of perceptual qualities</td>
<td>Brain (Later stages of neural processing)</td>
<td>Representations of reconstructed world</td>
</tr>
</tbody>
</table>

The first prism breaks the white light into violet, blue, green, yellow, orange and red, which are reconstructed into white light by the second prism. Likewise, our sense organs break the world into auditory, visual, tactile and olfactory components. When we use information from two or more senses at the same time, we reconstruct the world. There is no guarantee that the light entering the first prism is identical to the light that exits the second prism. In the diagram above, the dotted lines show that some wavelengths were reflected rather than refracted by the first prism, and thus, were not recombined by the second prism.

Likewise, when we break the world into perceptual parts we may miss things, for instance, information that sonar might provide or that an electron microscope gives us.\(^{11}\)

There is no guarantee that our reconstructed world is like the world that we encounter.

---

\(^{11}\) We also lose part of the range of each sense. For instance, our eyes are sensitive to light but our range excludes ultraviolet and infrared, our ears are sensitive to sound, but the range excludes sounds that other animals are sensitive to. The range for each sense varies for individuals, and are affected by disease, abuse, deterioration due to age, and conditions such as color blindness. Nevertheless, the range for each sense limits our information about the world in a different way. Extending the prism metaphor, this limitation would be represented as secondary reflection when the light exits the first prism.
through our senses. But this is beside the point. The problem under consideration was based on the fear that we would be unable to make a distinction between perceptual qualities and the objects of perception. And while it is true that if we consider only one sense, we cannot make this distinction, if we take two or more senses into consideration (as we must in dealing with all vertebrates), there is no reason why we should not be able to make this distinction in a meaningful though inexact way.

When two or more senses can be engaged, a vertebrate has a capacity to distinguish between the represented object and perceptual qualities. Indeed, one of the most predominant early instincts in human children (and other young vertebrates) is to engage two or more senses at once: they reach out to touch what they see, they turn to look at what they hear, they feel or look at what they taste (and put in their mouths what they touch or see). Thus, the object of a complex neuro-representation can be neither the state of our sensory organs nor the sensible qualities of external objects. We interact with the world through different senses and the resulting neuro-representations represent the objects that we sense in those different ways.

\[12\] Whether or not our reconstructed world is similar to the world which we live in doesn't matter. The reconstructed reality is the only one we can talk about or think about. And it is, for any purposes we can imagine, a fairly adequate world. We are also able to recapture some of the lost experience (i.e., catch some of the light reflected by that first prism) by creating scientific instruments that supplement our external senses: electron microscopes, sonar, infrared light detectors, etc. This increases the likelihood that our reconstructed world is similar to the one we live in (without, however, providing a guarantee). That perhaps is why science yields only theories.
Chapter 4

MEANING AND CONTENT:

FROM PUBLIC LANGUAGE TO PERSONAL MEANING

As a child grows and perceives the world around her, neural structures develop that represent the things that caused them. Some neural structures represent perceptual qualities — light, coloration, boundaries, shapes, sounds, softness or hunger. Other complex neural structures (which also incorporate simpler perceptual representations) represent the reconstructed objects of perception — colored objects, rolling balls, jingling bells, hunger and thirst-quenching bottles or hot stoves. These representational structures are produced by patterns of activity within the brain, and they affect subsequent patterns of activity, both within the brain and produced by the brain. When a person has a sensory perception, it may excite existent patterns of neural activity, thereby increasing the stability of the representational structure. By analogy, water flow creates channels in the ground, which affect subsequent flow of water. Water is then more likely to flow through the same channels, making them deeper, and thereby increasing the stability of those patterns of channels.

Representational neural structures are utilized in the production of overt behavior. Perhaps representations of hunger-quenching bottles and hot stoves are utilized in the production of behaviors that consist of reaching for things that look like bottles and shying away from stoves. When representational neuro-structures are utilized in the production of behavior, the resulting behavior is also representational (at least to some extent) and can, therefore, be interpreted.¹ When interpreting a representational structure, one

¹ Of course, we can also interpret non-representational structures if we take them to be representational, mistakenly or knowingly.
reasons from the representational content (discussed below) and one's knowledge of the medium of representation to the cause of the representational structure. Using knowledge of the human being as a medium of representation, we can infer from the behavior to the representational objects of the utilized neuro-structures. For instance, from the fact that a child shies away from the stove, we can infer that the child has a neural structure which represents the stove as hot or painful. In the past, one can reason, the child had experience with the stove which produced a representation of the stove as hot. The presence of this representation is evident from the child's behavior.

Learning involves the formation and utilization of representational structures. Learned behavior appears reasonable or meaningful since new behavior is interpretable in the light of the past experiences from which one learns. Not only do we infer past experiences from present behavior, but we explain present behavior on the basis of those past experiences. Given the child's past experience with stoves, it is only rational for her to shy away from them. The child learnt that they were hot in the past, and now avoids them. The child's behavior also seems to involve meaning — bottles mean food to her and stoves mean pain. Thus, she seeks the former and avoids the latter. Yet, although such meaning and rationality ultimately derives from the fact that the neural structures which make those behaviors possible are representational of past objects and events, the representational structures are not sufficient for rationality or meaning in this sense. The actions are meaningful only to an interpreter who infers from the representational aspects of behavior to events that must have occurred in the child's past. There is no reason to think that the child can interpret her own behavior, or even be aware that the behavior is interpretable. The child could be an automaton, merely utilizing the representational

---

2 Another child might not avoid them even though they are hot and this, too, may be rational given a different set of past experiences. For instance, she might have touched a stove on occasions in which it was not hot, and may be exploring where and when one can touch a stove without pain.
structures. For behavior to be meaningful or rational in this sense, all that is necessary is that representational structures are utilized in the production of behaviors. Any learnt behavior is rational or meaningful in so far as a knowledgeable interpreter can make sense of it, inferring from the behavior to the presence of representational structures and to the events that might have caused those representational structures.

There is a certain inevitability in the view that the behavior of people, vertebrates, some other animals and some computers is rational or meaningful in this minimal sense. Such beings can learn. They can develop representational structures that are later utilized in the production of behavior, so that the behavior is itself representational and thus can be interpreted. Hence, from the interpreter's perspective, the behavior is meaningful and in this way rational. Even linguistic behavior can be explained in this minimal way, for linguistic behavior utilizes representational neuro-structures. That is why linguistic meaning can be settled by the interpreter. Meaning, in this sense, lies in the act of interpretation rather than the acts of producing or using language. In this sense, words are meaningful to the interpreter, although we need to recognize that a speaker may be the interpreter of his/her own words.

Yet, some human activities, especially linguistic activities, seem to involve a more robust rationality and meaning than the minimal sense just outlined. It seems as if we mean things when we speak or act rationally prior to and independently of the interpretation of those acts. We don't just interpret our own words and those of others; we use words meaningfully. This intuition is reflected in phrases which imply that meaning is prior to the words or actions that convey it: choose your words carefully, say what you mean, can you put it into words, put it into action, or what did you mean by that? Phrases such as these suggest that meaning lies in or prior to the production of speech or other
acts, rather than merely in the interpretation of them. This recognition yields two further senses of meaning which emerge from my consideration of the acquisition and production of speech. First, there is a personal meaning (prior to speech acts) which stems from a person’s idiosyncratic experience with other language users; and second, an utterer’s or speaker’s meaning (in the speech act) which also links speech acts to other actions. The first I address in this chapter. The meaning found in speech acts I leave until chapter 6, when it can be discussed in the context of a general account of action.

Meaning and Content

Before addressing the question of personal meaning, it is important to discuss the relation between meaning and content. In the past twenty years, there has been much discussion of intentional content (also referred to as mental or propositional content). Discussion has focused on the content of propositional attitudes or thought attributions, and on the question whether such content is internal and narrow (or individualistic) or external and wide (or broad), or both. The current association between mental content and meaning stems from consideration of Putnam’s “The Meaning of ‘Meaning’ ” (Putnam 1975) in which he presented his notorious Twin Earth thought experiment. His point was, as he put it, “Cut the pie any way you like, “meanings” just ain’t in the head!” A few years later, Tyler Burge presented a new set of thought experiments quite similar to Putnam’s in “Individualism and the Mental” (Burge 1979). In contrast to Putnam’s focus on meaning, Burge emphasized mental content, by which he meant the content-clauses of mentalistic idioms such as “Tom believes that P” (that P' is the content-clause). Unlike Putnam who was primarily concerned with the effect of differences in physical environment on the meaning of natural kind terms (which, he argued, maintained their ‘meaning’ despite changes in our scientific theories), Burge was concerned with the effect of differences in a
person's socio-linguistic environment on the meaning of concepts and the individuation of mental content. He comments on Putnam's Twin Earth example:

Seen aright, the example suggests a picture in which the individuation of a given individual's mental contents depends partly on the nature (or what his fellows think to be the nature) of entities about which he or his fellows have de re beliefs. The identity of one's mental contents, states, and events is not independent of the nature of one's physical and social environment.

(Burge 1982: 107)

Burge's position was more general than Putnam's, as it applied not just to natural kind terms but to any concept; but it was also a much stronger claim. Not only meanings, but much of what we take to be mental is external (at least to some extent), since mental states, events or contents are individuated by reference to the external world, although mediated through consideration of the person's speech community. Despite the fact that this position went against virtually all traditional conceptions of the mind (which had tended to be internalist), it was extremely influential. Its rapid acceptance might perhaps be explained in terms of the widespread dissatisfaction in the decades since Ryle's The Concept of Mind with traditional conceptions of the mind. In addition, empirical psychology (profundly affected by behaviorism) and the growing success of neuroscience both suggested to many people that all human behavior (including linguistic behavior) could eventually be fully explained (at least in theory, if not in practice) without reference to any 'mental' or 'psychological' states. The idea that meanings were not individuated by consideration of a person's individual or psychological state (which does not lend itself well to 'objective' investigation, based on observations verifiable by multiple witnesses) was highly compatible with these trends towards the 'scientific' explanation of human behavior.

---

3 Since Putnam's and Burge's articles, most contemporary philosophers (including Fodor, Dretske, Davidson, Tye, Cummins, Chalmers, Bilgrami) concede that meaning and content must be external (wide), at least in part, (although there are some exceptions, for instance, Searle).
Yet, despite the promise of full explanations of human behavior found in the growing success of neuroscience, 'Folk Psychology' or 'belief-desire psychology' still offers better prediction and explanation of people's behavior. Folk psychological explanations of human behavior seek to explain what people do by reference to a wide range of propositional attitudes, not just beliefs, but also desires, fears, expectations, wishes, etc. Thus the content of a belief that it will rain is 'that it will rain', and the content of a wish that it will rain is the same. Given the general assumption that human behavior can, in theory, be given neuro-physiological explanations, one possible explanation for the success is of Folk Psychology is that there are neuro-physiological counterparts to propositional attitudes. However, for the purposes of explaining a particular person's behavior, it seems that narrow individuation of content is called for. Akeel Bilgrami gives a clear account of the conflict between narrow and wide conceptions of intentional content, which he claims have led to a 'bifurcation' of content:

Why should philosophers have turned against unity? Obviously, because they have felt a sense of tension between the two assumptions. This tension is usually brought out by way of examples. The general point of these examples is to show that if, as the second assumption says, content harbors elements that are external to agents, then intentional states will be impaired for the explanatory tasks demanded by the first assumption. The basic idea seems to be that since the object of explanations is to illuminate the behavior of individual agents, anything that is cited in these explanations must be constituted within these agents. The examples are supposed to establish this idea by making vivid the explanatory failures of those explanations which cite externally constituted contents. So, it is argued, one must posit a second notion of content, fumigated of these outlying elements which contaminate the externalist notion of content. This second notion will now carry out the explanatory tasks. (Bilgrami 1992:16)

Thus, discussions of mental content have become relevant to considerations of the status of Folk Psychology, whether the success of Folk Psychological explanations is due to the existence of neuro-physiological counterparts of propositional attitudes or whether it is merely a useful theory about human behavior.
Consideration of propositional attitudes also accounts for another connection between meaning and content. Although beliefs and other propositional attitudes need not be regarded as linguistic, they are expressed in words and are sometimes understood as dispositions to act in certain ways, including linguistic acts such as stating "It is raining." The content of a propositional attitude can be shared between speakers of different languages. John and Jean may have beliefs with the same content, even though John expresses his belief by uttering, "It is raining," and Jean expresses his own with the words, "Il pleut." Presumably the content of their beliefs is the same if and only if they mean the same thing when they say, "Il pleut" and "It is raining," which requires also that they have the same concept of rain.

Although the present spate of articles and books addressing the notion of mental content stems primarily from Burge's use of the phrase, a very similar notion of intentional content is found first in Gottlob Frege's article "On Sense and Reference". Frege writes, "We now inquire concerning the sense and reference for an entire declarative sentence. Such a sentence contains a thought." He then notes, "By a thought I understand not the subjective performance of thinking but its objective content, which is capable of being the common property of several thinkers." (Frege 1952:62)

In this chapter I want to return to Frege's conception of content. I do this for two reasons: first, I believe that internalist accounts of the mind and its content have advantages that have been overlooked in the recent rush to externalism; and second, I

---

4 It is in Frege that one first finds the idea that a 'thought' is the content of a declarative sentence, as well as the content of the content clause of sentences expressing propositional attitudes (of course, the term 'propositional attitude' comes from Russell). As I will discuss later, Frege did not consider 'thoughts' mental phenomena at all, distinguishing them at length from mental phenomena such as ideas, a general term he used to refer to mental experiences such as mental images, imaginings, perceptions, etc.

5 Frege is neither an externalist nor an internalist with respect to content; however, by identifying content with sense rather than reference, his position is anti-externalist.
want to use content to explicate meaning, and in more recent discussions, content and meaning are already so intertwined that the one cannot properly be used to elucidate the other. In Frege, though, the content (Inhalt) of a sentence (a thought) is distinguished from its meaning (Bedeutung), and identified with its sense (Sinn). 'Meaning' in this application can be understood as reference or denotation. The meaning or reference of names or expressions (components of a sentence) are the objects to which the names refer, and this is in keeping with the usage of contemporary externalists. Indeed, Max Black and Peter Geach translate 'Bedeuten' and 'Bedeutung' as 'stand for' and 'reference' respectively.

But we may set aside the difficult issue of how best to render Frege’s philosophical terminology into English. Frege certainly distinguishes between content and referential or extensional meaning. This, of course, leaves open the possibility that content itself (the sense of a sentence) is identified with another type of meaning such as intension. Frege gives quite a clear account of what he means by the content of sentences, and it is with this account that I start in developing an account of representational content, which I will then use in explicating personal meaning, and later speaker's meaning. I will return to the issue of externalism with a discussion of the Twin Earth thought experiments at the end of this chapter.

---

6 Frege, of course, argued that the meaning or reference of a sentence was the True or the False. Although few contemporary philosophers would assent to Frege's claim, some externalists do hold a position that can be profitably compared to Frege's: the reference of a sentence (or content clause) is a state of affairs which either is true or is false, or a state of affairs which is true in some possible worlds (true or false in our world), or set of satisfaction conditions, which either are or are not satisfied in the real world. See, for instance, David Lewis (1983), and also Milič (1984), although she uses a slightly different but related concept “real value” rather than reference.

7 Black and Geach note, "The natural rendering of these words would be 'mean' and 'meaning'; this rendering is actually required for their occurrence in German works quoted by Frege, and for his own use of the words when alluding to such quotations. But 'meaning' in ordinary English often answers to Frege's Sinn rather than Bedeutung. Russell's 'indicate' and 'indication' are barred because we need 'indicate' rather for andeuten. The renderings given here seem to be the simplest means of expressing Frege's thought faithfully. Philosophical technicalities, like 'referent' or 'denotation' or 'nominatum', would give a misleading impression of Frege's style." (Frege 1952: ix)
As we have seen, Frege did not consider a thought to be a mental phenomenon. This is, of course, also highly relevant to my discussion. Thoughts were, for Frege, neither part of the inner, mental world of conscious experience nor part of the external, physical world. Rather, thoughts and all other senses were part of the ‘Third Realm’, which is often compared to Plato’s world of forms. Many people, myself included, find Frege’s ontology problematic. Yet, Frege’s disturbing ontology is not a direct or inevitable consequence of his views on content and meaning. Thus, it is possible to put aside Frege’s ontology and consider only the properties he attributed to senses, concepts and contents of sentences; hence I will be offering an account of representational content which builds on Frege’s conception of content but does not commit us to his bizarre ontology.

When Frege first introduces the idea that the content of a sentence is a thought, he makes two important points. First, although acts of thinking are subjective, the contents of sentences (Fregean thoughts) are objective. Second, and closely related, the content of a sentence (the Fregean thought) can be a common property of different thinkers. His third point, which he makes subsequently, is that the content of the sentence is the sentence’s sense not its reference. Frege is motivated by his consideration of the sense and reference of the component expressions, words or signs (proper names) of a sentence. As he argued in the first part of his article, different words or expressions can have the same reference but different senses. Indeed, if one wants to make equality statements such as ‘the morning star is the evening star’ or ‘two plus two equals four’ meaningful, then it must be the case that the phrases on either side of the equality or identity sign must be the same in some way (same reference) yet different in some other way (different sense). Otherwise, we could not learn anything new from such statements, since they would be false (if there was nothing the same about the two expressions) or they would be
tautologous (since they would be equivalent to ‘the morning star is the morning star’, or ‘four equals four’). With respect to sentences, Frege writes:

Is this thought, now, to be regarded as its sense or its reference? Let us assume for the time being that the sentence has reference. If we now replace one word of the sentence by another having the same reference, but a different sense, this can have no bearing upon the reference of the sentence. Yet we can see that in such a case the thought changes; since e.g., the thought in the sentence ‘The morning star is a body illuminated by the Sun’ differs from that in the sentence ‘The evening star is a body illuminated by the Sun.’ Anybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false. The thought, accordingly, cannot be the reference of the sentence, but must rather be considered as the sense. (Frege 1952: 82)

The content of a sentence depends on the sense of its component words or expressions. Thus, the fact that a person might not be aware that two different expressions have the same reference implies that the content of a sentence produced by that person which contains such an expression cannot be individuated simply by considering reference.

Fregean contents, then, are objective, capable of being the common property of different thinkers, and cannot be individuated by reference alone. Frege also argues that contents, or senses in general, are neither part of the external, physical world nor part of the inner world of conscious experience. Although I do not agree that they occupy the ‘third Realm’, it is worthwhile considering Frege’s arguments that they are neither physical nor mental. In “The Thought: A Logical Inquiry,” Frege distinguishes between the common conception of the outer and the inner world.

A person who is still untouched by philosophy knows first of all things which he can see and touch, in short, perceive with the senses, such as trees, stones and houses, and he is convinced that another person equally can see and touch the same tree and the same stone which he himself sees and touches. ...

Even an unphilosophical person soon finds it necessary to recognize an inner world distinct from the outer world, a world of sense-impressions, of creations of his imagination, of sensations, of feelings and moods, a world of inclinations, wishes and decisions. For brevity I want to collect all these,
with the exception of decisions, under the word "idea". (Frege 1956:298-299)

Frege distinguishes between objects and ideas, or things in the outer world and things in the inner world. Objects in the outer world can be sensed, whereas ideas cannot be — ideas are had. "I see a green field, I have a visual impression of the green as well. I have it but I do not see it" (1956: 299) In addition, unlike objects in the outer world, ideas need a bearer. In order for things to exist in the inner world it is necessary for there to be a person whose inner world it is, "ideas need a bearer. Things in the outside world are however independent" (1956: 299). Ideas are, for Frege, sense impressions, feelings, imaginings, etc. and thus, they are fully subjective and thus individualistic — no two people can have the same idea. One only has access to one's own consciousness, and thus we cannot compare our ideas to ideas in other people's consciousness. Even if it was possible to transfer an idea from your consciousness to mine, Frege argues, since the idea would now be in my consciousness, I would have no way of knowing whether it was the same idea.

Thus, Frege concludes, "every idea has only one bearer; no two men have the same idea" (1956: 300).

Thoughts (the content of sentences), Frege argues, are neither ideas in the inner world nor objects in the outer world. Thoughts are not things in the outer world since they are not seen or touched, or otherwise sensed. People apprehend thoughts, or think them. Yet, like objects, and unlike ideas, one and the same thought can be apprehended or thought by different people. Otherwise, Frege argues, there could be no science. For if thoughts were purely subjective, then so too would be considerations of the truth-value of thoughts. Not only would my Pythagorean theorem differ from your Pythagorean theorem, but mine could be true while yours was false. "If every thought requires a bearer, to the contents of whose consciousness it belongs, then it would be a thought of this bearer only
and there would be no science common to many, on which many could work." (1956: 301)
Since we do not talk of 'my' Pythagorean theorem and 'yours', and since there manifestly are at least some common sciences, Fregean thoughts cannot be part of the subjective, inner world as are ideas (that is, sense impressions, memories, imaginings, etc.) Fregean thoughts, the content of propositional attitudes or sentences, are neither physical nor mental.

I assume that content should have the four characteristics that Frege attributed to sentential contents, or thoughts. First, they are objective. Second, content can be a common property. Third, content is sense rather than reference. And fourth, content is neither physical nor mental. The third and fourth characteristics imply that content of a representation can be identified neither with the represented object (the causes of the representation) nor with the interpreted object (that which one interprets as the cause of the representation). For the most part, represented and interpreted objects are physical objects, the things that interacted with the medium of representation thereby causing the representational structure on which we base our interpretation. These objects are also taken to be the reference of the representation: my father is the reference of a photograph of my father. (In developing an account of representational content, I begin by considering simpler, artistic representations such as paintings and photographs. I will then show how this conception of content also applies to representational neuro-structures.)

If we exclude the represented and interpreted objects, what other candidates are there for the content of a representation? The word "content" is, of course, a metaphor. Something (a sentence, a propositional attitude or a representation) contains something else. Thus, we can take this metaphor seriously with respect to representations and ask what do representations contain? Consider a visual representation such as a painting by
Constable or a photograph of my father. These representations cannot 'contain' the objects that cause them (the hillside and clouds that Constable painted or my father); nor can it 'contain' people’s interpretations of the picture or the photograph. But, these representations do contain images. I maintain that content should be identified with the image in a visual representation, and with the counterparts to images in other types of representation, including neuro-representations.

One might object that the visual representation is the image, and thus it cannot contain the image. But this is not right. I have been using representation as a short-form for 'representational structure'. The representational structure is much more than the image: it is also the paper, canvas, paint and pigment. An image is not identical to the material substance — the medium — in which it exists. After all, many prints of the same photograph contain the same image, as do the Constable painting and a reproduction of it (in a different medium altogether). Same image, different representations: image and representation are not identical. Nor are images identical to the objects that they represent — an image of my father is not my father. Nor is the image just that which people interpret it to be — it is an image of my father whether or not anyone recognizes it as an image of my father.

Understanding the content of a visual representation as the image can be seen to satisfy at least the last three of Frege’s conditions. The image can be a common property of different representations: different representations, different photographs or paintings, can contain the same image. The image is not the reference of the representation, that is, it is not identical to the external objects that caused (or can be interpreted as having caused) the representational structure. The image can be understood as the sense of the representation. Like Fregean senses, different images of the same object (different
photographs of my father) need not be recognized as having the same reference (photographs of the same person). Thus, it would be meaningful to say, 'these photos are of one and the same person', or 'this person [pointing to one photograph] is this person [pointing to the other].' In addition, the image is not a physical object — it is not the representational structure, made of canvas and paint or paper and pigment. Nor is it a mental object, since it is contained in the representation itself. It is best understood as the set of representational features in the medium, abstracted from the medium, or the set of alterations to the medium of representation, but abstracted from that medium. These features or alterations are abstracted in the sense that they are not dependent on that particular physical structure. For instance, literally, the alterations to a wax impression made by a coin are shapes in the wax. If one abstracts those shapes from the wax, then they are just shapes, and could just as easily be shapes in plastic or shapes in silly putty. Thus, different media of representation can share representational content.

The representational structure is a purely physical structure which can be understood ahistorically — it is of such and such a shape and size and depth, or consists of certain connections or weightings of those connections, etc. But, the content of a representation cannot be understood ahistorically — an understanding of something as a set of alterations to a medium must include some conception of a previous state of that medium whether or not one has any notion of what that state would consist of. Thus, the content of a representation can be understood as the alterations to the structure of the representation abstracted from the physical substance of the medium: the impression in the wax made by the ring, and thus not itself formed of wax. Extending this to neural representations, the content would be the patterns of connections caused by our sensory experience abstracted from the physical substance of the brain.
The content of a representation stands in the middle between representation and interpretation: it is the set of features in or alterations to the medium of representation, caused by interaction with the represented object, and on which interpretations of the representation are based. Thus, the content of representation satisfies Frege's first condition in that it is objective — not dependent on the subjective experience of those interpreting the representational structure. The content of a representation is determined by the interaction between the medium and the represented object (the impressing of the wax by the signet ring). Yet, in acts of interpretation the content of the representation is also dependent on the observer in an important way. For it is the observer who must abstract the content from the substance of the medium. The content of the photographic image is not the stuff of which the photograph is made (the photographic paper, etc.), but the image in the photograph. The content of the wax seal is not the wax itself, but the impression in the wax. The content is constituted by the structural features that are attributed to the represented object rather than to the medium (for instance, in the wax seal, the shape of the impression may be attributed to the represented object, but the color may be attributed to the medium). Such structural features are what an observer reasons from when interpreting the representation. Thus the representational structure, which is independent of observers (but entails the possibility of an observer), constrains but does not determine the representational content. The final determination of the content rests with the observer. To this degree and in this sense the content is observer-dependent.

Although content, as I use the word, is in this way relative to an observer, there is also a sense in which a representational structure has an ideal content. The ideal content would be the complete set of alterations or structural features caused by interaction between the represented object and the medium of representation. When different people
look at a representational structure, the same physical structure which they observe constrains their interpretations. They base their interpretations on different content, and yet, since they are looking at the same structure which constrains their content, it is desirable to be able to talk about the content which they all partially grasp. This generates the artificial concept, "ideal content" — an indefinite, possibly infinite, set of structural features, which cannot be grasped by any individual. The ideal content would be the complete set of features on which an 'ideal observer' could base an interpretation — the ideal observer being that impossible, godlike creature, with unlimited and unerring observational powers (including senses of observation not shared by humans, such as echolocation, etc.) and perfect knowledge of the properties of the medium and with the capacity to observe the representational structure from all possible points of view. The concept of ideal content is meant to capture the full set of observer-independent constraints imposed by the representational structure, while simultaneously recognizing that content in any act of interpretation is observer-dependent, and thus necessarily incomplete.

Obviously, there is no guarantee that a represented object and the object which we interpret a structure as representing will be identical. Sometimes, we misinterpret representations — we make mistakes. These mistakes may be quite reasonable. After all, the content of the representation on which our interpretations are based is often, by itself, insufficient to determine the identity of the represented object. For instance, when I look at a picture of one of the twins, the structural features on which I base my interpretation could have been caused by either Colin or Cole. The content is indeterminate between an image of Cole and an image of Colin. When I look at the picture, I may interpret it as a picture of Cole, utterly unaware that he has a brother. Such mistakes may be unnoticeable,
unverifiable, unfalsifiable, or irrelevant to the human practices which involve those very representational structures. For instance, if I am showing the picture to another person so that he can meet Cole (whom he has never seen) at the airport, this mistake (since it is actually a photo of Colin) will be irrelevant. Thus, the relation between the content of the representation and erroneously interpreted objects may be as important to us, in some contexts, as the relation between the representational content and the ‘truly’ represented object.

For the same reasons, the representational content can serve as a link between objects that are in no other way related. A reasonably but erroneously interpreted object will be similar to the represented object in so far as they would produce similar representational structures in the relevant medium so that they could both be reasonably inferred from the representational structure. For instance, two signet rings cast from the same mold but of different metals (one gold, one brass) may leave quite similar impressions in sealing wax. Since the content of the representation is observer-dependent, if an observer can observe no difference between the impression of one ring and the impression of the other, then the content of the two representational structures is the same with respect to that observer. From the point of view of the observer, the wax structure could have been produced by either ring, and thus, is conceptually (but not causally) linked to either object. In this way, the representation may also be conceptually linked to objects that do not but could exist, for instance, any other object that could leave the same impression. The objects that could produce a representation with the same content could be quite different from one another. A ring, a button, a ceramic press, a wooden block, etc. may all be engraved or molded such that they could all leave the same impression in wax.
Thus, the content of a representation can act to create natural, but observer-dependent, categories. These representational categories are the sets of objects that are conceptually linked in that they could cause in the medium the structural changes on which the observer bases his/her interpretation. These categories are observer-dependent since they are determined by the structural features that the observer notices in the representational structure and takes to be representational in that he/she reasons from those structural features when interpreting the representational structure. Depending on the observer's observation skills, familiarity with the medium, his/her access to the representational structure, etc. an observer will attribute a certain content to a representational structure. An indefinite set of objects could produce representations in that medium with that content. This set is what I call a representational category.

Although representational categories are observer-dependent, they are also natural in that 1) the representational structure itself is caused independently of the observer (as discussed in the previous chapter) and 2) given the content, the set of objects that could have created a representational structure in that medium with that content is determined by factors independent of observation. For instance, consider animal tracks. When I look at animal tracks, the content that I abstract is quite general, four holes of an approximate size and shape. I do not notice details in their spacing and size. Indeed, the structural features of the tracks from which I reason when I interpret them are so crude that either a cat, a rabbit or a large squirrel could have left tracks with such content. In contrast, when an expert looks at the same tracks, she reasons from much more specific content, four indentations with more specific shapes, sizes and spacing. Indeed, the only naturally occurring objects that could leave tracks with these structural features are middle-sized rabbits.
The composition of such natural, but observer-dependent categories will depend on many factors. The more skilled and more familiar with the medium that the observer is, the smaller the category is likely to be. The more sensitive the medium of representation is and thus, the more detailed the representational structure is, the smaller the category is likely to be. Two people looking at the same representational structure, but who have different observation and interpretation skills and techniques, may observe and/or reason from different structural features, thus observing it to have different content. Thus the natural category associated with each content would likely differ. On the other hand, two people with similar experience and observation skills may reason from the same structural features, and thus, for them, the representational structure to have the same content, then the same category will be created for both. The full extension of a representational category, since it contains non-existent entities, is indefinite and cannot be fully known. Given this fact, as well as the fact that media are not perfectly sensitive to each and every structural feature of a represented object, even ideal content is associated with a representational category which could be infinite in size.

In social contexts, individual differences in what is taken to be the content are minimized, and thus, the related categories tend to converge. For instance, if the animal track expert points out the structural features of the tracks on which she bases her interpretation, then I too begin to consider these features as representational and thus expand what I take as representational content. If a type of representational structure (such animal tracks) is important in a certain context, people are likely to spend more time observing such representational structures; hence aboriginal hunters have more stable representational categories in this area than urban office workers do. The more time and experience one has with a medium of representation, and representational structures within that medium, the more structural features one is likely to observe and reason from.
Thus, as people become more experienced with a medium of representation, they are likely to become aware of more representational features. As observational skills, knowledge of the medium, and experience increase, their individual content increases and may even begin to approach the ideal content (the complete set of structural changes caused by the interaction between the represented object and the medium). Thus, given that human beings begin with the same observational equipment (eyes, ears, etc.), differences in content may diminish as people gain experience, even in the absence of communication between observers.

When a representational structure is utilized (rather than interpreted), only the representational content (and not the object of representation) is relevant. When we make a print from a negative, the content of the negative is transferred from the negative to the print. Whether the negative represents Cole, Colin, Cordelia, a house or the forest, etc. is irrelevant to what the new print will look like. Provided the new medium is sensitive to structural features in the negative (the content) then the print will have similar content, and represent the same object as did the negative. In contrast, what was standing in front of the shutter was not irrelevant to the formation of the first representation — the image on the film. If something else had been standing there, it would represent something else. (This would be true even if the content were to remain the same.)

Yet, despite the fact that only content (and not the represented object) plays a role in simple utilization, this is not always the case for interpretation. When we interpret the photograph, we do not base our interpretation merely on the content. We also utilize knowledge of the medium and context, as well as other relevant facts and beliefs. Exactly how we utilize these things depends on our interests, our skills, and the context. Thus, despite the fact that the content of the photographic representation is utterly indeterminate
between Cole and Colin, we can sometimes determine who is actually represented. For instance, we may be able to trace the history of the photo, and determine that it was taken at a time and place which coincides with Cole's presence, but precludes Colin's. Even when we cannot fully determine the identity of the represented object, our interpretations are often partially based on information that is independent of the representational content, even in contexts where, ultimately, it doesn't matter what the represented object is. We use such information even more in contexts in which it matters who actually stood in front of the camera, despite the fact that the photo looks as much like Colin as it looks like Cole. If Cordelia wants to know what her estranged father looks like, then perhaps either photo would do. But if she wants a keepsake of her father, it may matter to her that it is a photo of him rather than of his brother. In such a situation, Cordelia's knowledge that her Aunt Clara gave her the photo, and said that it was taken when her father visited Clara in England, can provide additional information that allows her to infer from the representation to the represented object — her father. In interpretation, although we reason from the content of the representation, we can also draw on other information in our reasoning.

By beginning with a Fregean conception of content, we were able to develop an account of representational content that can be used to explicate different conceptions of meaning, personal meaning (discussed in the next section), speaker's meaning (discussed in chapter 6), as well as the meaning attributed to a speaker in acts of interpretation. To summarize this notion of representational content: the content of a representation is the set of alterations to the medium of representation, abstracted from that physical medium, that were caused by the represented object. A representation represents that which was causally or historically responsible for the structure. When we interpret a representation, we reason from the structure to its cause. But the content of the representation, which is the basis of our interpretation, is not to be identified with that which 'causes' it. Rather, it is
the set of features caused by the represented object's interaction with the medium, which
the interpreter abstracts from the representational structure when interpreting. For the
most part, only the content of a representation is relevant to the utilization of a
representation in the production of an interpretable, or meaningful, product (which would
also be representational). However, in acts of interpretation, the content is not the only
relevant route to the represented object.

In neural representation, structures are produced in the brain as a result of neuronal
activity associated with the people's interaction with the sensible world. In theory (but not
in practice) these structures might be interpreted by a neuroscientist. If a neuroscientist
could interpret such representations, she would reason from the content — the structural
features which she attributes to the represented object, or objects, not to the prior structure
of the brain. Associated with the content would be an indefinite category of objects, or sets
of objects, that could have produced representations with the same content in that person.
As the skill of the neuroscientist increases, or as neuroscience itself develops, the content
expands or becomes more specific, and the associated category narrows.

Of course, it is likely that no one will ever be able to interpret representational
neuro-structures — the neuro-scientist sketched above is as much a part of science fiction
as Twin Earth itself. However, we can and do interpret representational behaviors which
are produced through the utilization of representational neuro-structures. When I sit in the
philosophy department lounge, I see numerous people go up to the vending machine, put
in four quarters or a loonie (a dollar coin) into a slot, press a button and receive a can of
pop. When people do things like this, we can reasonably infer that they have had past
experiences with vending machines, coins and cans of pop. Indeed, in order for us to
interpret their behavior in the way that we normally do (they are purchasing pop from a
vending machine), we must assume that they have had past experiences that they have learnt from, experiences which caused neuro-representations of vending machines, coins, etc. that are utilized in their present behavior. The representational content of their behavior seems to be pretty much the same. That is, if I abstract the representational content of their behaviors from the particular movements of their particular bodies, they do the same thing (purchase a can of pop from a vending machine).

Yet although they do 'the same thing' it is unlikely that these behaviors are caused by identical brain structures. Thus, one might question how it is that their behaviors could all have the same content given that the brain structures that those behaviors depend on are likely quite different from one another. We have seen that different objects can cause representations with the same content; for instance, different signet rings can cause the same impressions in wax, in some plastics, in playdough, etc. When the medium theory of representation is applied to the brain, this implies that different experiences can cause neural representations with the same content. Every year the Canadian Mint produces thousands of quarters that are all pretty much alike. Each of these quarters could be represented by wax impressions which are so alike one another that no one could tell which coin was the cause of which. Although each impression would represent the quarter which was pressed into it, the impressions would be alike in content. Any of these impressions could have been produced by a different coin (or even by a non-coin, such as a hard plastic 'quarter'). Likewise different people's neuro-representations may have the same content even though the representations are caused by different experiences. For any neuro-representation, there is a representational category that consists of the set of experiences that could have caused a representational category with the same content. Since different sets of experiences can produce neuro-representations with the same
content, different people can have neuro-representations with the same content as mine if the sets of experiences which cause the representational structure are members of the representational category associated with the content of my neuro-representation. If my neuro-representation of vending machines was caused by my experiences with vending machines, but other experiences would have been sufficient to cause a neuro-representation in me with the same content, then why should not other people's experiences with vending machines cause neuro-representations in them with the same content as my neuro-representation?

When I watch people purchase pop from a vending machine, I take for granted that they have all learnt what vending machines are and how to use them, that is, I take for granted that their past experiences with vending machines have caused alterations in their brain structures and in their behaviors. When I consider what they have learnt, how they have been altered, I conclude that they have all come to represent the machine as a vending machine, the metal disks as coins, the cans as cans of pop, etc. I infer that they have all had past experience with vending machines and coins, but although I have a pretty good idea of what sorts of experiences they are likely to have had in the past, I have no idea as to what particular experiences any of these people had. An infinite number of different sets of experiences with vending machines and coins could produce the same representational behaviors. In explaining their actions as buying a can of pop from the vending machine, it doesn't matter at all which particular experiences any of these people have had. The content of their behavior (with respect to vending machines) is the same. I can therefore interpret their behavior as utilizing the representations of coins, vending machines, etc. that are also alike in content, despite the fact that they were caused by different experiences.
The content of a person's behavior is the set of changes to a person's pattern of behavior, abstracted from the behavior in acts of interpretation, that are taken to be caused by past experience. The child now shies away from hot stoves (and we assume that she did not always do so) because in the past she had experiences with hot stoves from which she learned. Thus, her behavior can be interpreted as meaningful or rational. In the next section, I show how this general theory of representational content can be applied to speech behavior.

Personal Meaning

Personal meaning can be understood as the representational content of linguistic neuro-representational structures — the set of alterations to the medium of the brain caused by sensory experience of people using words. One can also understand personal meaning as the representational content of speech behaviors — the changes to a person's behavior caused by their past experiences, particularly of people using words to communicate. Using the conception of content developed above, this implies that personal meaning is determined internally (if considering the representational content of linguistic neuro-structures) or individualistically (if considering the representational content of linguistic behaviors). But how do we acquire linguistic representational neuro-structures and how do we use them? In other words, how do we learn a language?

The account of language that I present is in many ways similar to that of Locke. On Locke's account, words are signs for ideas or concepts. Words are the means by which we communicate our ideas to others as well as the marks with which we express our ideas to ourselves for use within our minds. Thought is, for Locke, prior to language use, for unless we have ideas, we cannot use words to communicate these ideas, even if we
have the physical capacity to produce the sounds of the words. We learn to attach words that we hear other people speak to our own, individualistic ideas. Thus, we use public words to express personal ideas or meanings, even if our own ideas or meanings are very poor representations of the external objects that cause them. Locke writes:

2. The use men have of these marks being either to record their own thoughts, for the assistance of their own memory; or, as it were, to bring out their ideas, and lay them before the view of others: words, in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them, how imperfectly soever or carelessly those ideas are collected from the things which they are supposed to represent. ... A man cannot make his words the signs either of qualities in things, or of conceptions in the mind of another, whereof he has none in his own. Till he has some ideas of his own, he cannot suppose them to correspond with the conceptions of another man; nor can he use any signs for them: for thus they would be the signs of he knows not what, which is in truth to be the signs of nothing. But when he represents to himself other men's ideas by some of his own, if he consent to give them the same names that other men do, it is still his own ideas; to ideas that he has, and not to ideas that he has not.

3. This is so necessary in the use of language, that in this respect the knowing and the ignorant, the learned and the unlearned, use the words they speak (with any meaning) all alike. They, in every man's mouth, stand for the ideas he has, and which he would express by them.

(Locke 1959:9-10)

On Locke's account, words are signs for ideas, or mental representations. The account of representation that I have developed is, in contrast, an account of physical representational structures, that is structures in physical media. Thus, although the account of language I will present is very much like Locke's, it does not depend on a mental medium of representation, but merely on the physical medium of the brain. On my account, words are not used to express ideas or concepts; rather, words express the representational content of neuro-structures. Yet, the functions of language remain the same — to record one's own thoughts (through thinking in words or talking aloud) for the aid of memory (as well as in verbal activities such as reasoning and deliberation) and to
bring out one's thoughts and communicate them to others. Whereas Locke held that one must first have ideas before using words to stand for them, I hold that one must first have representational neuro-structures before one can use words that express their content, since this activity of using words expresses the content of neuro-structures by utilizing those neuro-structures. Although different people in a linguistic community use the same words when speaking, each person uses the words to express his own particular neuro-representational content, since, after all, one can only utilize one's own particular neuro-structures in the production of speech (or any other) behavior.

Since people use words to express their own individualistic concepts and ideas, this implies that inexperienced language users and master language users are doing the same things with their words, and their success and failure to use words 'properly' by the standards of their linguistic community will be measured in the same way. No matter how poor our neuro-representations are (how poorly the structures represent the objects that caused them) the words associated with them function in the same way to express the contents of those representations. The person who uses a word incorrectly by the standards of his linguistic community (using 'maple tree' to refer to a sweetgum tree) is doing exactly the same thing as the person who uses it correctly — utilizing his own representational neuro-structures in the production of public speech. Thus, the central question in the discussion of language concerns the relations between personal meaning and public language.

If all learning involves the formation and utilization of representational structures, then learning a language must begin with the development of representational structures specific to language use. The earliest quasi-linguistic neural representations would be the representations of word sounds, or in the case of congenitally deaf children, word signs or
mouth movements. There is no need to posit anything unusual about these representational structures. They would be simple perceptual neural representations, caused by the activity in the brain when the sense of hearing is stimulated by spoken words, or other word-like sounds. The child’s hearing of repetitions of certain sounds, for instance, the word sound “Mama” or “Cat” would eventually result in stable neural patterns, or perceptual representations of those word sounds. I call such representations “quasi-linguistic” because they are representations of the word sounds,⁸ and not of words. For these sound representations would not yet be associated with other representations, for instance of people or animals. In general, they would have no linguistic or semantic associations or functions.

As discussed in chapter three, the objects of primary perceptual representations are perceptual qualities (colors, sounds, shapes, etc.) rather than complex external events or objects. Likewise, these quasi-linguistic representations would represent the sounds of words. Yet, perceptual representations also become connected to one another in different ways, as a result of co-activation. Repeated simultaneous activation of different neural perceptual representations results in associations between representations and in complex representations. As discussed earlier, the objects of these representations are the reconstructed objects of perception. Thus, if visions of plush toys co-occur with feelings of softness, then not only will perceptual representations result (representations of the shapes and textures of plush toys and of feelings of softness), but a complex representation will also be produced. The object of the complex representation is the reconstructed object of perception, the toy that looks plush and feels soft.

---

⁸ I use ‘word sound’ rather than ‘phoneme’ since it is more inclusive. Of course word sounds consist of ‘phonemes’ and a child probably learns to represent both the individual phonemes and the sequences of phonemes that constitute word sounds.
Since complex representations are caused by the co-activation of neural representations, complex representations may also incorporate representations of word sounds which may frequently co-occur with representations of objects. If the word sound "Teddy" co-occurs with other perceptions of a stuffed teddy bear, then a complex representation would result. Like other complex representations, it would represent a reconstructed object of perception — the toy bear which has a certain appearance and feel and which is called "Teddy". Thus, the word sound "Teddy" becomes an integrated part of the reconstructed object of perception. Unlike Locke's account, in which a word is attached to the idea that the word expresses, on my account, the word is often an integral part of the neuro-physiological counterpart to the idea — a complex neuro-representation.

Children who grow up in a community of language users will have neural representations of the world such that objects in the world are fully integrated with the words that we use to talk about them. Many of even the earliest complex neural representations would be composed of tactile, visual, auditory, other perceptual and quasi-linguistic components. Thus, the representation of the word sound should not be thought of as connected to or associated with the representation of an object of perception. Rather, the objects of perception are themselves quasi-verbal, since they are constructed from linguistic, visual, auditory, tactile and other perceptual components. Indeed, ‘word’ might be profitably compared to observed properties of objects (such as coloration, shape, sound, and texture) which comprise the object of perception, rather than be understood as labels attached to, associated with, or standing for objects. Just as we experience objects as colored when we look at them, so too we experience objects as ‘worded’ or having a name. For the most part we see an object and its coloration in one and the same act; and likewise, we see an object and see what it is at one and the same time, whether or not we
are actively noticing its coloration or what it is. Of course, the perceived color of an object changes in different situations, in different lighting conditions and for different people; however, no matter what coloration we experience when we look at an object, we experience it as colored. Likewise, the word associated with an object changes from one linguistic community to another; however, no matter what words we use, if we are language users, then we experience objects as 'worded' or as something we can talk about.

When a child learns her first language (in normal circumstances, in which the child grows up among language users), many of the early, stable complex representations are likely to be integrated quasi-linguistic representations. The representations are in constant flux, especially when the child is young, since stability of representations is generally the result of repeated exposure to similar stimuli. Stable representations result from repeated overlapping of patterns of activity. Although there is a tendency for previously activated patterns to become re-activated as a whole (which will speed up the stabilization process,) until a child has a lot of experience with things, neural representations will be quite unstable. For stable complex representations to develop, there must be frequent co-occurrence of different perceptual representations.

This means that early exposure to language may affect the ways in which complex representations stabilize. First, hearing words used would increase the amount of co-occurrence between perceptual representations since it increases the possibilities for co-variance. Not only will there be co-variance between visions, sounds, tastes, smell and textures, but there will also be co-variance with word sounds. More importantly, even before a child learns language, hearing word sounds can help the child learn to categorize objects and events into the very groups that common words pick out. With any complex representation, activation of one part of the neural representational structure can result in
the activation of connected parts. For example, consider a child who has developed a stable representation that integrates the vision of dogs, with the smell, feel and sound of dogs, and the word sound 'dog'. On occasions on which a dog is merely seen, the entire complex representation may be activated, even though only the visual representation is directly activated, and may, in fact, be activated to a greater extent. Likewise, when the child hears the word 'dog' the entire complex may be activated. Now, suppose the child sees a dog. Like any perceptual experience, that causes neural activity, including activation of visual perceptual representations. Suppose, in addition, that since this dog happens to be much bigger than any other dog she has seen, the complex dog representation is not activated. If, however, the child hears the word sound 'dog' at the same time, it can result in the activation of the complex dog representation, causing new connections between the complex dog representation and the patterns of activity that the sighting of a large dog did excite. Just as seeing animals that trigger the dog representation increases its stability and scope, so too does hearing the word sound 'dog' in the presence of animals which otherwise would not trigger the dog representation.

As the child hears people call more dogs 'dogs', her neural representation of dogs will change. As it changes, the range of animals which trigger the complex dog-word representation will also change, but these changes will be in accordance with how the word is used in the child's linguistic community. Thus the representational categories associated with these dog representations will also be in keeping with the category of dog within her linguistic community. This effect of hearing language on the child's categorization scheme is pre-linguistic — the child is not a language user, nor can she be said to understand the language. Yet when the child later becomes a language user, she will have a great
advantage. Her neural representations will already be functioning so as to sort the world into the same categories that words happen to pick out.

First language learning would differ in this way from foreign language learning. When learning a second language, one forms connections between the stable complex representations (which integrate objects and first-language word sounds) and new, less stable foreign-word perceptual representations. In contrast to first language learning, in which the word-perception complex representation develops and stabilizes as a unity, in foreign language learning a relatively unstable word representation must be incorporated into a fully formed complex representation. True fluency is not achieved until the new words are fully integrated. Until then, speaking a foreign language involves connecting the foreign word to a fully formed native word-object complex. Indeed, this is the type of model that Locke presents for first language learning.

Children who do not hear or see language used early in their lives must have few complex representations in which quasi-linguistic and other perceptual representations are fully integrated. Thus, for these children the process of learning language would be more like foreign language acquisition. They would presumably already have some form of stable representations of the reconstructed objects of perception. They would then have to form stable representations of words sounds, and form connections to the more stable, complex representations of objects, eventually forming new complex representations that integrate word and object. However, unlike children who grew up amidst language users, the representational categories of such language-deprived children may be quite unique. For instance, they might not recognize different species of dogs or individual dogs as belonging to a single category that corresponds to our word, 'dog'. Although the fact that humans are built pretty much alike might keep these categories from being wildly inconsistent, the inconsistencies that would arise could dramatically impede language
acquisition. The same principles should also account for the occasional failure of translation between different languages (linguistic communities may carve up the world differently), as well as the general success (people are built and behave pretty much alike).

The convergence between a child's representational categories and those of her speech community begins long before the child becomes a language user, and continues throughout the process of language acquisition. But in order to become a language user, the child must also learn to produce word sounds. At some point the child begins to babble, and to make wordlike sounds. If she produces wordlike sounds in public contexts, she will sometimes elicit responses from other people. For instance, others may interpret (rightly or wrongly) the wordlike sounds as triggered by a nearby object, and thus, as an attempt to talk.

Whether or not the child's production of sound is in fact an attempt to talk, other people may respond to her as if it were. They may give her the object that it seems to refer to, repeat the word for the object more clearly, say a different word, etc. Responses which focus attention on the object and/or a word for the object will excite the complex representational structures, and begin to build stable connections between them and the neural structures involved in the production of word sounds. As connections stabilize between the neural structures involved in producing sounds and the neural structures that represent objects, the child's first word sounds are refined until they sound more or less like other people's. Gradually utterances will be more often triggered by perceptions of the objects with which the word sounds tend to co-vary.

Some productions of sound may be triggered by perceptions. For instance, when seeing a dog, the child may produce the word sound 'dawg'. The production of word sounds in a young child may in this way utilize neuro-structures which represent the
reconstructed, quasi-linguistic world of the child's experience. On seeing a stuffed bear, the child may produce the word sound, 'Teddy', utilizing that representational structure which integrates the word sound 'Teddy' with other perceptions of the stuffed animal. Such speech behavior is also representational, since these speech behaviors are produced through the utilization of a representational structure. Like the representational structures utilized in the production of the speech behavior, her words represent the reconstructed object of perception — the stuffed bear which has a certain look and feel and is called 'Teddy'. Her utterance, triggered by the perception of the bear, has an inherent meaning. Even prior to anyone's interpretation, the child's utterance of 'Teddy' represents the bear which she perceives and has heard called 'Teddy'.

The meaning of a child's first words is to a large extent personal. The word, after all, represents the reconstructed bear of the child's experience: reconstructed from countless past perceptions of the bear and of it being called "Teddy". Likewise, when the child uses the word 'dawg' in the presence of a dog, she is drawing on the representational structure produced by her own experience with dogs and doglike animals. This is the same representational structure that failed to be activated by a large dog, since she had had no experience with such dogs. At any given time, a person's quasi-linguistic representation of an object will have incorporated only those uses of a word that he or she has experienced. Especially in the early years, this subset will be quite idiosyncratic due to the child's relative lack of experience.

The content of a neuro-representation is the set of alterations produced in the medium of the brain by a person's interaction with her environment. Associated with any representation is the category of objects which might also have caused a representational structure with the same content. The child's dog representation happened to be caused by
her own particular experiences with dogs. But there is no reason why another set of experiences could not have caused the same representational neural structure. The possible set of experiences that could have caused such a representational structure is quite constrained by the brain structure and perceptual apparatus, and the fact that the connections are built from countless experiences. On the other hand, there is no reason to suppose that the child’s representational structure would have been significantly different if she hadn’t seen the particular dogs that she did see, or if she hadn’t seen them on those particular occasions. Thus, the associated representational category is the set of dogs or doglike things that could have caused a representational structure with the same content, and which would be, for all practical purposes, any set of dogs and dog-experiences similar to the ones that the child actually had.

Closely connected to the associated representational category is the range of objects that would trigger the activation of the representational structure. Any perception that can trigger the representational structure must have a significant amount of overlapping neural activity. Significant overlapping neural activity and the activation of the representational structure also alter the representational structures, increasing their stability and/or the connections between them and other representational structures. Thus, the kind of perception that can trigger the activation of a representational structure is precisely the kind of perception that could cause the representational structure. If it wasn’t the kind of perception that could cause the complex representational structure, then there wouldn’t be enough overlap in neural activity to trigger it. Thus, if one wants to know what kinds of animals produced in a child a representation of dogs, the easiest way to find out is to see what kinds of animals the child calls ‘dogs’.

The representational structure represents the set of objects that caused it, that is, all those dogs that she perceived and/or heard called ‘dog’. Also associated with the
representational structure is the representational category — those objects that could have caused a representational structure with the same content. The representational category would be almost identical to the category of things that would trigger activation of the representational structure. For the most part that representational category would include dogs, things that look like dogs and things that are called ‘dogs’. Unfortunately, just as some things that look like dogs but aren’t dogs will trigger the representational structure, some perceptions of non-dogs probably helped to cause the representational structure. This might appear to be a problem since the dog representation will have been caused by, and thus represent, a few (if not many) cats and bears along with all the dogs.

This is, of course, a version of what Fodor calls the ‘disjunction problem’. If either a cat or a dog can trigger the representational structure, then surely it is a cat-or-dog representation rather than a dog representation. Any naturalizing theory of content is susceptible to the disjunction problem. If content of a representation is identified with its cause (through co-variance or functional role) then there isn’t room for failure. As Fodor puts it:

You have to get error in somewhere, and so far we’ve made no room for it. In fact, there looks to be a dilemma about this. Suppose, to put it crudely, that ‘dog’ means dog (and thus has dogs and only dogs in its extension) because it’s a law that dogs cause ‘dogs’. (Fodor 1980: 59)

If only dogs cause ‘dogs’ then nobody makes mistakes, and if sometimes non-dogs (for instance, a cat on a dark night) cause ‘dogs’ then ‘dogs’ must mean dogs-or-cats-on-dark-nights and not dogs. Either way, there is no room for error. And, making matters worse, not all dogs cause ‘dogs’. In addition, there can be difficulties accounting for the relation between personal meaning and the social or public nature of language. Even when a person uses the word ‘dog’ because dogs cause ‘dogs’ for him, he may simultaneously intend to use ‘dog’ in the same way that others in his speech community use ‘dog’. The set
of things that cause 'dogs' for the individual may not match up with the set of things that cause 'dogs' for others in his speech community.

In fact, in so far as the personal meaning of word is concerned, the representation that integrates perceptions of a cats and dogs with the word sound 'dog' is a cat-or-dog representation rather than a dog representation, at least from the point of view of others in the speech community. Marika, when she was just over one, said 'dawg' whenever she saw dogs, cats, and any other mid-sized, furry quadruped. From her behavior, we can infer that Marika had a complex representational structure which integrated the word sound 'dawg' with other perceptions of animals. The associated representational category was mid-sized mammals rather than dogs. From the perspective of a competent language user, the representational structure that integrated the word sound 'dog' with other perceptions was a cat-or-dog-or-rabbit-or-other-mid-sized-animal representation. Yet hearing the word sound 'dog' in the presence of dogs is supposed to produce representational categories that are in keeping with the linguistic categories of the community. Marika's category was clearly out of sync.

From the point of view of a competent speaker of English, Marika is making a mistake. She wrongly utters the word 'dawg' around mid-sized mammals that are not dogs. Indeed, this is quite a common mistake for children to make in my linguistic community, although some children utter 'cat' rather than 'dog' in much the same circumstances. On this model, this type of error is common and, to a large extent, corrigeble. Any complex representation is due to a repeated co-activation of perceptual representations. For instance, Marika can develop a complex representation of 'objects-that-look-soft' if tactile perceptions of softness tend to co-vary with visual perceptions of plushness. Things that look soft feel soft, so Marika develops a representation of things that look and feel soft.
This representation is utilized in her behavior — she reaches out to pet things that look soft. If she reaches out to pet one of those cacti that look soft, then she will discover that she is mistaken. But this is by no means the end of the story. This experience will likely cause significant changes to her representational neural structures such that the associated category of things that look and feel soft does not include those cacti. (Actually, it may take a few encounters with the cacti before enough restructuring takes place that her behavior towards such cacti changes.) My cat, perhaps, makes this type of error when she comes running at the sound of a can opener when I am opening a can of tomatoes. But even her representational neural structures can undergo change, and she eventually learns to stay put, except at certain times of day.

The same type of error occurs when learning language, and it is corrected in the same way. If a child’s representational categories (integrating perceptions and word sounds) do not match up with linguistic categories, then this becomes apparent in public contexts. Although Marika charmed cat and dog owners alike when she uttered the word ‘dawg’ on seeing their pets, there was no doubt that she was using the word incorrectly and they responded accordingly. Some responses (for instance, pointing at the cat and saying ‘No, it’s a cat!’) might have resulted in Marika increasing the amount of time she spent looking at the animals, which may result in more discriminate visual representational structures. These and other responses probably also resulted in development and stabilization of other neural structures, for instance, one that might represent the word sound ‘cat’. But, often when Marika uttered word sounds in public, more restructuring occurred. Marika eventually learnt to sort cats and dogs into different categories, and learnt to call some of them cats. Such mistakes, or inconsistencies between personal representational categories and linguistic categories, will occur throughout her life, but as
she grows up in a community of language users, they will occur less and less often. When
she gets really good at sorting animals, she may even be able to identify ones that don't
belong in any of her categories. (I don't know what kind of animal that is!)

Unfortunately, even as Marika develops into a competent language user the
disjunction problem lingers. After all, even though the representational structures continue
to change and the child gradually learns to utter words appropriate to the circumstances, it
still remains the case that some of the perceptions which caused a dog representational
structure were perceptions of cats. Thus, no matter how good Marika gets at calling only
dogs 'dog', the neural structure utilized in this behavior will represent a few cats along with
the dogs.

The final solution to the disjunction problem is two-fold. The first part consists in
acceptance. Some quasi-linguistic representational structures probably represent things
that the linguistic community generally wouldn't accept as part of the extension of the
relevant word. For instance, I suspect that I have often mistaken grackles for starlings. I
have only recently developed a representation of grackles, and prior to this year, every
grackle I saw, if I categorized it at all, I would have categorized incorrectly. Thus, I suspect
that some of the bird perceptions that caused my starling representation (which dates back
to my childhood) were, in fact, perceptions of what others in my speech community would
call 'grackles'. (Of course, I cannot be certain, since it never occurred to me that the birds I
saw weren't starlings which is why they reinforced the starling representation.) Not only did
perceptions of grackles activate and cause my starling representation, but quite likely
perceptions of grackles still would, whether or not they also triggered my grackle
representation. In such a case I might say, "There's a starling... no, I think it's a grackle."
The vast majority of us have a wide range of disjunctive representations of this type —
think of all the trees, birds and weeds you confuse with one another. With respect to the *personal* meaning of the words we use, some representations just are disjunctive (when compared to the 'dictionary' or 'expert' meaning of the word).

On the other hand, even disjunctive representations can cease to be disjunctive. In general, the greater the number of similar things or experiences which activate and thus cause a neural structure, the more stable the representational structure will become. If stability is caused by the overlapping of neural activity, then the changes to the structure caused by aberrant activity could eventually be eradicated by future activity. By analogy, suppose the previous owner of my house planted a garden. Over the years some plants die, I plant some new ones, I change the borders, cut down a tree, etc. Both the previous owner and myself are causes of the garden's structure. But, eventually, none of the original garden exists. At that point, I am the sole cause of the garden. Likewise, even if there are originally multiple causes of a neural structure, eventually the structure may change so that some of those original 'causes' are no longer causally relevant.

Suppose Marika's 'dog' representation was originally a cat-or-dog representation, caused by the neural activity involved in seeing dogs and cats and hearing some of them called 'dog'. When she utters the word 'dog' in the presence of cats, people tend to respond, "No, that's a cat". Thus, a new representational structure begins to develop that integrates the word sound 'cat' with perceptions of cats. Sometimes both representational structures are triggered by perceptions of animals, but gradually, social forces result in cats triggering 'cat' representations more often than they trigger 'dog' representations. Thus, 'dog' representations get triggered more exclusively by dogs. This reinforces patterns of activity that are more exclusively triggered by dog perceptions. Eventually, the neural patterns that were originally triggered by cats and dogs are eradicated. They fade.
the child begins to call only dogs 'dog' we can assume that the parts of the representational structure that were caused by cats are no longer causally relevant to the representation. Like the garden whose structure was eventually entirely caused by me, it is theoretically possible that, at some point, the integrated 'dog' representation could be entirely caused by dogs. The personal meaning of the word 'dog', at that point, would no longer be disjunctive. Yet it is unlikely that many (if any) of us ever develop 'pure' or non-disjunctive concepts, such that we could never be mistaken in our use of the word. However, for the purposes of adequate communication, it is necessary only that concepts can become less and less disjunctive with respect to our linguistic community as we gain experience with the public uses of words.

As people are socialized, the personal meaning of their words becomes more and more consistent with the meaning within the linguistic community. The process of assimilation begins long before language use, from the very first time a child hears language. The process of assimilation is greatest when the child is learning to speak, but continues through one's entire life. Many failures of correspondence probably go unnoticed, since in most contexts it doesn't matter if I mean something slightly different from you, provided that we agree that the word applies to the present situation. Some failures of correspondence may account for shifts in the public use of words. Hearing a person use a word in a slightly unusual way could cause changes to other people's linguistic representations, and eventually cause them to use it that way as well. But often, when meanings fail to correspond, it gets noticed. These are errors in usage with respect to the linguistic community. And when errors get noticed, they often get corrected, resulting in further convergence between personal meaning and public use.

But, the meanings of people's words will always have a personal aspect. The process of developing complex representations that integrate words with perceptions
ensures that our personal meanings will be to a large extent consistent with public meanings. After all, these representations depend on co-variance between word sounds and perceptions — co-variance which is determined by the linguistic community. But the same process also ensures that our personal meanings will fail to correspond exactly, since we can only experience a small fraction of the uses of words by our linguistic communities.

The process of learning language and the process of communication through language are one and the same. Communication through words consists in utilization of neuro-representations in the production of speech behavior. These neuro-representations were formed, and are continually being altered in small ways, through our experience with other people's use of words. When we hear other people speak, these neuro-representations are activated, resulting in understanding (to the extent that we already have and utilize such representational structures) and learning (to the extent that every activation causes some alteration to the structure and its connections to other representational structures). In this way we can communicate 'ideas' from one person to another. Communication of ideas or thoughts is bound to be imperfect since the personal meanings connected to the words with which we express our ideas are idiosyncratic; yet the communication of ideas succeeds to a large extent because we have developed these personal meanings in much the same environment. Marvin Minsky makes a similar point about communication and the learning of language:

A word can only serve to indicate that someone else may have a valuable idea — that is, some useful structure to be built inside the mind. Each new word only plants a seed: to make it grow, a listener's mind must find a way to build inside itself some structure that appears to work like the one in the mind from which it was "learned." (Minsky 1986:270)
Twin Earth Considerations

Personal meaning, the content of integrated linguistic neuro-representations, is determined solely on the basis of a person's neurophysiology, or on the basis of the linguistic behavior that the neurophysiology produces. Such meaning is individualistic; determined by a person's idiosyncratic brain structures and thus, literally, 'in the head'. In contrast, Putnam's Twin Earth thought experiment and Burge's similar thought experiments are used to show that meaning is external. In this final section, I discuss the relation between Twin Earth considerations and this internalist conception of personal meaning, or representational content, and the implications for the 'meaning' of meaning.

According to Putnam's original thought experiment, Oscar₁ and Oscar₂ are neurophysiological and psychological duplicates, in two different worlds, Earth and Twin Earth respectively. These worlds are identical in all ways but one: water on Twin Earth is not H₂O, but a different chemical, abbreviated 'XYZ'. The year is 1750, and thus, nobody in either world knows that water is H₂O or XYZ. Putnam claims:

Oscar₁ and Oscar₂ understood the term "water" differently in 1750 although they were in the same psychological state, and although, given the state of science at the time, it would have taken their scientific communities about fifty years to discover that they understood the term "water" differently. Thus the extension of the term "water" (and, in fact, its "meaning" in the intuitive preanalytical usage of the term) is not a function of the speaker by itself. (Putnam 1975: 224)

Putnam is motivated by his concern that the meaning of a natural kind term (for example, 'water') should stay constant through changes (even radical changes) in scientific theories about those substances which form the extension of those terms (for example, water). If the extension or meaning of a term is fixed by a person's psychological state then the extension of the terms which we use to talk about substances will often change when we have new scientific theories about those substances. Yet, a new theory about the nature
of what we call 'water' cannot be new theory of what we called 'water' prior to the theory

*unless the meaning of 'water' stays constant through the theory change.*

In "The 'Meaning' of Meaning", Putnam contrasts his own position with that of an antirealist:

To a strongly antirealist intuition it makes little sense to say that what is in the extension of Archimedes's term \( \chi\rho\upsilon\sigma\delta \zeta \) is to be determined using our theory. For the antirealist does not see our theory and Archimedes's theory as two approximately correct descriptions of some fixed realm of theory-independent entities, and he tends to be skeptical about the idea of 'convergence' in science — he does not think our theory is a better description of the same entities that Archimedes was describing. But if our theory is *just* our theory, then to use it in deciding whether or not \( X \) lies in the extension of \( \chi\rho\upsilon\sigma\delta\zeta \) is just as arbitrary as using Neanderthal theory to decide whether or not \( X \) lies in the extension of \( \chi\rho\upsilon\sigma\delta\zeta \). The only theory that is not arbitrary to use is the one the speaker himself subscribes to.

(Putnam 1975:236)

Putnam argues convincingly that since extension and truth are closely tied, the antirealist ought to reject the notion of extension, as he rejected that of truth. The realist, on the other hand, must recognize (as Putnam did) that extension cannot be a determined solely by a person’s psychological state (or by intension or sense). Archimedes meant by \( '\chi\rho\upsilon\sigma\delta\zeta' \) what we mean by 'gold'; even if it turns out that some of the things that Archimedes was warranted in calling \( '\chi\rho\upsilon\sigma\delta\zeta' \) by his theory are things that we are not warranted in calling 'gold' by ours. If our theory is right with respect to whether or not those items fall into the extension of gold then, with respect to those items, Archimedes was just wrong.

The Twin Earth thought experiment makes this point even more strongly by showing that two people, Oscar\(_1\) and Oscar\(_2\), can be in the same psychological and neurophysiological state while understanding the term 'water' differently; since 'water' for Oscar\(_1\),
means H₂O and 'water' for Oscar₁ means XYZ. They mean different things by their terms even though, in 1750, they would be equally warranted in calling the same samples 'water'.

Putnam, of course, is correct that the stuff that the two speakers are talking about when they use the term 'water' is different: Oscar₁ is talking about stuff that scientists will later understand to be H₂O and Oscar₂ is talking about stuff that scientists will later understand to be XYZ. This is what Putnam calls "extension" or "meaning" in the "intuitive preanalytical usage of the term". On Putnam's account, the meaning of a natural kind term remains constant throughout time, no matter how much our scientific theories change.

In 1750, a substance was picked out by the term 'water'. The term 'water' continues to pick out exactly that stuff — to mean that stuff — as scientific theories about its nature change. Whether or not we have the capacity to determine if a sample liquid is a sample of that stuff is irrelevant to the meaning of the word 'water'.

The term 'meaning', as Putnam allows, is used in many different ways. Putting aside considerations of whether or not there is an "intuitive preanalytical usage" of the term 'meaning', and whether or not Putnam is right that this notion of extension captures that "intuitive preanalytical usage"; it appears that the term 'extension' is also used in many different ways. Putnam seems to be using extension to pick out what I have called the 'represented object'. The represented object is, again, that which is causally responsible for the representational structure in the medium. Oscar₁ and Oscar₂ are causally connected to different liquids, and whatever representational structures they utilize when they use the word 'water' (or indeed, engage in any water behavior such as swimming or drinking it) represent different objects. Using extension to mean 'represented object', Putnam is clearly correct when he says that the extension is not a function of the speaker alone. The represented object remains the same through time, through changes in
scientific theories and from one context to another. As I discussed in chapter two, the represented object is determined by the interaction between a medium of representation and objects in the world — our ability or inability to interpret correctly it is irrelevant to what the representational structure represents.

Yet, there are also other notions of extension. Burge, for instance, identifies extension, not with the represented object, but with the interpreted object, which (as we saw earlier) need not be the same. Whereas the represented object is causally connected to a person, the interpreted object is that which interpreters infer (from a person’s speech behavior) to be causally connected to that person.

The differences between Earth and Twin Earth will affect the attributions of propositional attitudes to inhabitants of the two planets, including Adam and Adam*. The differences are not to be assimilated to differences in the extensions of indexical expressions with the same constant, linguistic meaning. For the relevant terms are not indexical. The differences, rather, involve the constant context-free interpretation of the terms. (Burge 1982:107)

This interpretational notion of extension is the basis of many social theories of meaning; although, of course, even in such theories, the social relations often mediate causal or historical relations with objects or substances.

Yet, there is a third notion of extension that emerges from consideration of personal meaning and its relation to representational content. This notion of extension is best captured by the ‘associated representational category’ — that is the set of objects that could have caused a representational structure with the same representational content. Despite the fact that a particular object or set of objects actually caused the representational structure in a medium, and thus are represented by it, other objects or sets of objects could have caused a representation with the same content. If extension is understood as the associated representational category, then extension, in this sense, is solely a function of the speaker. Using extension in this sense, both H₂O and XYZ are in
the extension the term ‘water’ when either Oscar₁ or Oscar₂ uses the word. Although the object of Oscar₁’s representation of water is H₂O and the object of Oscar₂’s representation of water is XYZ, both Oscars have representations of water with the same content (they have the same concept of water), and the content is such that both H₂O and XYZ are included in its extension. The concept ‘water’ is, as Frege suggested, a function which moves from sense (or content) to extension.

Yet, using this conception of extension, it is not immediately clear that the extension of the term ‘water’ has the realist quality that Putnam regarded as central — ‘water’ means that stuff that I drink, that I point at and that I have learnt to call water. If the extension includes both H₂O and XYZ (and who knows what else), then ‘water’ couldn’t possibly mean (in the sense of extension) that stuff that I have learnt to call water since I have only learnt to call H₂O ‘water’. It would seem that this is exactly the conception of extension that Putnam is criticizing. In the remaining portion of this chapter, I will show that we can meet Putnam’s realist concerns without rejecting an internalist conception of personal meaning and the related conception of extension.

On Putnam’s view, the natural kind terms (and some other terms) have an indexical quality; their extension is determined through act(s) of ostensive definition, such as ‘baptisms’. Searle, however, argued against Putnam that meaning need not be external even if indexicality is central to such terms, which is doubtful in many cases: “Most people do not go around baptizing natural kinds.” (Searle 1983:208) Searle’s argument turns on the possibility that the indexical definition of the term can itself be ‘in the head’. He writes:

(1) The associated cluster of concepts does not determine extension,

(2) The indexical definition does determine extension,

(3) What is in the head does not determine extension.

Now (3) does not follow from (1) and (2). To suppose that it does one must assume that the indexical definition is not in the head. Putnam
uses (1) and (2) to argue for (3) and thereby assumes that the indexical definition is not in the head. (Searle 1983:205)

Searle argues that although the lexical meaning of an indexical expression is not by itself sufficient to fix the extension, the lexical definition provides a rule for determining reference relative to any utterance of that indexical expression. Applying this to natural kind terms, the sense of a word such as ‘water’ would be insufficient to pick out its referent by itself; however, the sense would also provide a rule for determining reference relative to any utterance of ‘water’. Thus, one could, in theory, preserve the ‘eternal’ quality that Putnam associated with natural kind terms. Searle does not explain how it would work for natural kind terms, but presumably it would work something like this: ‘water’ would mean H₂O since anytime (even on Twin Earth) an ordinary Earthling uses the word ‘water’ (in an ordinary way) he means H₂O, because the sense of his term ‘water’ prescribes that when he uses the term he means that stuff that the term ‘water’ was used to pick out as he acquired the term ‘water’. In other words, his term ‘water’ has the sense: whatever is identical to this stuff.

Searle is right that some relation between the utterance and the term uttered must be central to fixing the extension of natural kind terms if they are understood as having an indexical component, as well as to fixing the extension of ‘pure’ indexicals, such as ‘you’ or ‘now’. However, there are problems with attaching any rule for determining reference to the sense of individual terms. Although, I do think that such an approach might be warranted for pure indexicals (such as ‘you’ or ‘now’), for which the sense is no more than a procedure for determining reference — learning how to use words such as ‘you’ or ‘now’ is just learning how to use them to refer to different people or times in different contexts. ‘I’ is an exception in so far as it is tempting to think of it as having a more robust sense since, for example, it seems that I can only use the word ‘I’ to refer to Niko Eugenia Scharer.
Thus, it might seem that it has, for me, much the same sense as ‘Niko Eugenia Scharer’. Yet, given that we understand without any problem other people’s use of ‘I’, and occasionally use ‘I’ to refer to other people (when quoting them or putting words in their mouths), it is more reasonable to treat it like any other indexical in so far as its only sense is a procedure for determining reference, but that procedure determines that in ordinary circumstances one can only use the word ‘I’ to refer to oneself.

Searle’s claim that natural kind terms could also have an indexical component, such that the sense of the natural kind terms defines a rule for determining extension, is problematic. This ‘rule’ would have to be something above and beyond the criteria for picking out samples of the substance associated with the term, criteria that are determined by the intension of the term. The ‘rule’ would have to be something of the nature: when this term is uttered it picks out the same stuff as was picked out by certain past utterances of the word (for instance, the utterances from which one acquired the term, utterances that were baptisms or utterances in which one used it successfully). There are, of course, going to be problems determining which past utterances are the relevant ones. Even more problematic is that this kind of ‘rule’ isn’t really indexical at all. An indexical expression has a sense which is self-referential to the present utterance of the expression. Yet, in the case of natural kind terms, one need not consider the present utterance or context of use at all. If anything, certain past utterances or contexts should be considered. Searle can be seen as extending the ‘sense’ of such terms beyond the cluster of concepts to include the following caveat: the extension of this term is fixed by my causal or learning history, or by a ‘baptism’. The problem is that, although the rule is ‘in the head’, the causal or learning history certainly isn’t. Thus, extension is ultimately fixed by external events.
Although Searle's solution is problematic, he is right to hold that even if natural kind terms have an indexical component, meanings need not be deemed external. What would it mean to say that a natural kind term functions as an indexical? As an indexical, the contexts in which we use the term should determine the extension. When I say that I will meet with you tomorrow, the context in which I utter this (say, on October 5, 1998) determines what day I am talking about (October 6, 1998). When I say that I would like a glass of water, the context in which I utter this (say, in your kitchen) determines what substance I am talking about (the stuff that comes out of your tap in the kitchen, or the stuff in the Brita® container on the counter). In most contexts, there is no ambiguity. Water means H₂O in most Earthly contexts, in some contexts which involve drinking it means relatively pure H₂O, in some contexts which involve swimming it means relatively large amounts of not-necessarily-potable H₂O, in some contexts which involve oceans it means saline H₂O, etc.

This indexical component is not part of the sense of the term when 'sense' is taken to mean the content of the linguistic neuro-representation — there is no rule associated with the sense of every term that states that the occasion for use determines extension. The indexical component applies to any use of the term, and the varied ways in which we use a term depend on our linguistic neuro-representations. Thus, it seems that even on my account, in which the content of the linguistic neuro-representation of water is indeterminate between H₂O and XYZ, in so far as there is an indexical component to our use of terms such as 'water' it is reasonable to say that for us, water means H₂O, since whenever we use the term 'water' the context makes it clear that that stuff is what we mean. This remains true even when scientific theories of water change, and thus, does not depend on our knowing the underlying or hidden nature of water. The indexical nature of
natural kind terms, which comes into play on occasions on which we use a term, links the content of a linguistic neuro-representation (which is ambiguous between all members of the associated representational category) to the represented object.

There are, of course, hypothetical contexts of use which might appear to be problematic. For instance, suppose I say that I would like a glass of water, but it turns out that I woke up this morning on Twin Earth rather than my native Earth. In one sense, the context of use is Twin Earth where ‘water’ means XYZ, thus the context suggests that I mean XYZ. After all, any Twin Earthling’s utterance in the same context would undoubtedly mean XYZ. In another sense, the context of use is pretty much the same as every other time I said that I would like a glass of water, and thus, the context suggests that I mean H2O.

What these examples really show is the extent to which we must make assumptions of stability when using words meaningfully and when interpreting others who use words meaningfully. Using words meaningfully (in the ways that we have always used them) generally requires that we have made a Humean assumption that things are much as they were before\(^9\). The sudden switch to Twin Earth threatens this assumption, and thus, interferes with the ordinary functioning of natural kind terms. Likewise, problems with the meaning of ‘water’ would arise if things in the world keep changing, for instance, if the underlying nature of water keeps switching — H2O today, XYZ tomorrow, C2H5OH the next day, etc. One could not claim, in such a world, that natural kind terms have any sort of indexical component. Indeed, one might even have trouble with ‘pure’ indexicals and demonstratives, such as ‘here’, ‘now’ and ‘this’.

When I am transported to Twin Earth, I might still assume that nothing has changed, and thus, I mean H2O just as I always have meant it in such contexts. My new

\(^9\) We can make this assumption explicitly or implicitly through our use.
acquaintances, Twin Earthlings all, also assume nothing has changed, and thus, they interpret me as meaning XYZ just as my twin always did. On the other hand, if I know or suspect that something has changed (that I have left my home planet), then I probably am a little more careful when I say, 'I want a glass of water'. I can no longer assume that everything here is just like it has always been on my native planet (although it looks that way). Yet, if I still assume that nothing has really changed back on Earth, then when I do use the word 'water' I still mean H₂O, just as I always have.

Content, personal meaning, is indeterminate between all members of the representational category associated with a neuro-representation. Yet, despite this, when we use a term, the context in which we use it often determines the represented object. In this way it functions like an indexical. But, for indexicals (and perhaps all meaningful utterances), we must make certain assumptions about the world when we use such terms meaningfully, i.e. that it remains stable. Oscar, and Oscar₂ have the same representational content. It is ambiguous between H₂O and XYZ. Yet, on most occasions on which either of them use the term 'water', the context links the term to the represented object of the linguistic neuro-representation — H₂O for the Earthling, XYZ for the Twin Earthling. Thus, although extension can be identified with the representational content, on most occasions in which we use a term, the context serves to link the term to its represented object (or extension in Putnam’s sense).

The position I advocate is quite similar to, but not identical, to one that Putnam rejects: the term ‘water’ is world relative but has a constant relative meaning. Thus, ‘water’ has the same meaning on Earth and on Twin Earth; yet, water is H₂O on Earth but XYZ on Twin Earth. On my account, Oscar, and Oscar₂ have the same personal meaning for the term ‘water’ in so far as the representational content of their integrated word-object ‘water'
representation could have been caused either by $\text{H}_2\text{O}$ or by $\text{XYZ}$. The associated representational categories are the same for each. Yet, I would not, for merely this reason, claim that the personal meaning is exactly the same for Oscar, as it is for Oscar$_2$. The medium theory of representation is a causal theory — representations are *history-laden* structures. The fact that Oscar$_1$'s water representation was caused by his interaction with $\text{H}_2\text{O}$ whereas Oscar$_2$'s water representation was caused by his interaction with $\text{XYZ}$ is also relevant to the personal meaning. Indeed, it is precisely because Oscar$_1$'s water representation was caused by his interaction with $\text{H}_2\text{O}$ (Oscar lives on Earth where water is $\text{H}_2\text{O}$) that any context in which Oscar$_1$ uses the word 'water' is sufficient to determine that the extension is $\text{H}_2\text{O}$.

Thus, the term 'water' on my account is not world relative. It is the case that the representational content, personal meaning, of the two Oscars' terms are the same in so far as either substance could have caused a representational structure with the same content. In this way, the associated representational categories are the same. However, representations are history-laden structures; whenever Oscar$_1$ uses the word, 'water', he utilizes neuro-structures that were altered by his past interaction with $\text{H}_2\text{O}$, whether or not he (or any other member of his community) is aware that water is $\text{H}_2\text{O}$. Thus, any context in which uses the word 'water', the basic Humean assumption that the stuff that he is now calling 'water' is the same as the stuff he learnt to call 'water' fixes the extension of his term to $\text{H}_2\text{O}$. Thus, the sense of extension given by representational content meets Putnam's

---

10 There are, of course, times when this assumption is false; for instance, when Oscar calls a glass of gin, 'water'. This is not a problematic failing of this assumption; since it does not threaten the meaning of water (see the discussion of disjunctive concepts in the previous section). However, if the assumption is false because the underlying nature of water has suddenly changed (for instance, from $\text{H}_2\text{O}$ to $\text{XYZ}$ or some other substance), then this would indeed cause problems. As I indicated earlier, if the nature of substances regularly changed in such a way, then we would probably have to abandon the notion of extension altogether, as well as give up on indexicals and demonstratives.
realist concerns in any context of use that doesn't threaten extension altogether (such as a context in which the nature of water doesn't remain constant).

Thus we can compare Twin Earth scenarios to photos of twins. We have two photographs in front of us, and each looks as much like Colin as it looks like Cole. Indeed, for many purposes, either photograph can function just as well as a photograph of either twin. After all, had either photograph been taken of the other twin, it would look identical. When we utilize the photographs without interpreting it (for instance, making a copy of it), only the content matters. But when we interpret the photograph, then much more than the content comes into play. In particular, the ordinary interpreter knows something of photography and knows that one twin or the other stood in front of the lens when the button was pressed, and that each photograph is 'really' a photo of that twin. Whether or not it is in our interests to try to determine which twin is in the photo, the fact that we know that it is a photo of one or the other can affect how we interpret it and how we use it in interpretive contexts.

Similarly, in Twin Earth scenarios, we postulate two people, twins (one on Earth and one on Twin Earth) who have identical (or similar) brain structures. Each brain structure was caused as a result of interaction between the twin and the substance (water or 'twater') found on his planet, prior to the development of molecular theory. These brain structures are alike in content (both ideal content and any individual observer relative content, since no observer should be able to tell the difference). Thus, the related representational category would (for both contents) include water, 'twater' as well as other hypothetical or even actual substances (perhaps heavy water). In this way, each brain structure seems to represent water just as much as 'twater'. Yet we know something of neurophysiology and we know that each is 'really' a representation of the stuff that the person came in contact with. Still, in some circumstances either brain state can function
just as well as a representation of either stuff. (After all, had the person been causally connected to a different stuff, their brain state would be identical.) When these brain structures are merely utilized in the production of certain behaviors (for instance, drinking it, splashing it), only the content of the neural representation is relevant. The twins' neuro-representations will be utilized in the same ways, producing the same behaviors. However, when we interpret those neural representations, or when we interpret behavior that was produced through the utilization of such neural representations, then other things become relevant, especially our knowledge that there is a represented object that is causally connected to the content, that an indefinitely large set of experiences could produce representations with the same content, as well as other knowledge we may have of the link between the represented object and the representation.

The content of the twin Oscars' neural representations may be identical and thus function identically in certain contexts. However, in the context that we are in when considering the twins' neural representations, and interpreting the twins' behavior which is directly caused by those neural representations, it matters that we have other sources to draw on in our interpretations. We know, for example, that two different molecular substances are involved, and that each twins' neural representation was causally connected to one twin or the other. We even know which is connected to which. This was all given as part of the thought experiment. We consider this information when we interpret the twins' behavior. Thus, we can say, as Putnam does, that Oscar, means H_2O and Oscar_2 means XYZ.

The twins, on the other hand, aren't interpreting their neural representations. They aren't even necessarily interpreting the behaviors that utilize these representations (drinking, swimming or talking about water). But what happens if the twins and their
colleagues do engage in interpretation? They will be able to interpret (correctly) the
represented object of their behavior (or their neural representations) as the stuff that they
have always been in contact with. They can also correctly interpret the content of their
representational behavior as related to many substances that would produce in them a
representation with that content. That is, recognizing that their behavior with respect to
water depends on neuro-representations formed when they had past experience with
water, they might also recognize that other substances (water, 'twater', heavy water, etc.)
could have caused neuro-representations with the same content (in so far as it would
produce exactly the same behavior). As far as their subsequent behavior is concerned, it
may not matter at all which substances happened to produce neuro-representations with
that content. Different substances could have produced the same content. Indeed, given
the advance of chemistry on Twin Earth, and the development of chemical theories quite
similar to our own, they may even interpret, *incorrectly*, but reasonably, that the
represented object as H₂O when in fact it is XYZ. (Likewise, even here on Earth, water
may turn out to be something other than H₂O at least according to future scientific theories.
But nevertheless our neuro-representations of water will still represent what we now call
'H₂O'.)

Although the content of a representation underdetermines the identity of the
represented object, the representation should not be considered ambiguous on those
grounds. To begin with, all representations would be ambiguous in this way, since there is
a category of potential causes associated with every representational content (because the
category is not limited to existent objects). But, more importantly, the photograph does
represent one twin or the other (whoever was standing there in front of the lens).¹¹

¹¹ Of course, some pictures really are ambiguous. The duck-rabbit picture represents whatever it is
causally and historically connected to. Since it is reproduced in philosophy and art books precisely
because it is ambiguous and looks like both a duck and a rabbit, in those instances, it is causally
Likewise, each of the twins' neural representations represents either water or 'twater' (whichever the twin grew up around). Of course, the content of the representation would be the same for either, and likewise the associated representational category as well as the set of things that are likely to cause water behaviors. However, the represented object would not be the same for the two twins.

The Twin Earth scenario cannot show that 'meaning ain't in the head' any more than a photograph of a twin can show that photo albums don't contain images of people. Photographs are representational. The content of the photographic image is in the photograph, but is abstracted from it in acts of interpretation. The represented object, the photographed twin, cannot be the same thing as the image that he is the cause of. It is because the photographic medium is sensitive to visible objects that the photograph can be either be correctly interpreted as representing the person that stood in front of the camera or be erroneously interpreted as representing someone else. Likewise, it is because the vertebrate brain, when located in a sentient being, is a medium sensitive to the sensible world, that people can be interpreted as representing the world that they sense. The medium's capacity to represent objects ensures that some interpretations (within the representational category associated with the content) will be reasonable for any given observer, and that some interpretations will be correct, but it does not ensure that they will be reasonably or correctly interpreted. But we can misinterpret what those things stand for only because they in fact do stand for things in the first place. The neural representations which we utilize in behavior make those behaviors meaningful — these meanings are in the head.

c connected to ducks and rabbits, and so represents both. But, no doubt at some point, somebody actually drew a representation of a rabbit and found it was easily interpreted as a duck. When I was a young child I made an owl out of clay. It represented an owl, but was consistently interpreted as a cat. If I drew it now, I would do so because it can be interpreted as both an owl and a cat, and thus it would, in this context, represent both.
Putnam, in "The 'Meaning' of Meaning", allows that there are many components of meaning, all of which he takes to be external. Although much of his attention is focused on extension, he also considers related senses of meaning such as predominant and deviant senses (for instance, lemons picked from trees are lemons in the predominant sense, whereas plastic lemons are lemons in the deviant sense) and stereotypes. In this respect, that there are many components to meaning, I think Putnam is right. Yet, there are more components than Putnam allows. Consider first the concept of extension. We have isolated three notions of extension which contribute to meaning: the extension in the sense of represented object (Putnam's sense), the extension in the sense of the interpreted object (Burge's sense), and the extension in the sense of the category associated with the representational content (my sense). This third sense of 'extension' is internal; although it meets Putnam's realist concerns with respect to the effect of theory change on extension. Likewise, the closely related personal meaning is determined by the psychological or internal state. Thus, in addition to the many components of meaning which admittedly are not 'in the head' (represented object, interpreted object, stereotypes, etc.) two central components of meaning, namely personal meaning and the extension which is fixed by the personal meaning, are fully in the head.
PART II
MENTAL AND OVERT BEHAVIOR AND ACTION
Chapter 5

MENTAL BEHAVIOR

Children don't just learn to speak with words, they learn to think with words. Verbal thought is an almost universal mental behavior. It can be profitably compared to other 'first-order' behaviors such as speaking aloud, running or eating. Such behaviors are first-order in that they can be thought of as the primary skills on which more complex behaviors and practices are based. For instance, linguistic behaviors and practices such as promising, preaching, verbal humor and comedy, 'dissing' and swearing all involve overt speech and cannot be understood except within the context of a language-using community. For the most part, you cannot perform one of these behaviors unless you can talk aloud (or use sign language), but talking aloud is neither necessary nor sufficient for their performance. Of course, people seldom engage in a first-order behavior without also doing something more complex. For the most part, if you are speaking you are also doing something else — perhaps promising, telling a story, giving orders, preaching or swearing. Similarly, other first-order behaviors such as running, throwing, or walking are skills which comprise or are incorporated in complex behaviors and actions, and are rarely (if ever) performed alone. This is true also of verbal thought. Countless mental behaviors are forms of verbal thought or incorporate verbal thought — silent prayer, deliberation, mental rehearsal, mental arithmetic, silent reading, psyching oneself up, etc. Verbal thought is one of the most prevalent first-order mental behaviors, and thus, I will use it as my central example.

In the last chapter, I dealt extensively with the formation of representational neuro-structures and the utilization of those structures in the production of overt linguistic behavior. If thinking in words is a mental behavior, akin to the overt behavior of speaking
aloud, then the behavior should be produced in much the same way: sensory stimulation causes patterns of brain activity which are determined by neuro-structures, producing further patterns of brain activity that are involved in the production of speech. However, unlike the brain activity of the overt behavior, that involved in mental behavior does not result in the movement of muscles and audible language. I will treat verbal thought as a type of linguistic behavior in which the overt signs are actively suppressed. Although the observable movements are suppressed, the conscious experience that also accompanies most overt linguistic behavior remains, allowing the language user to experience his or her own mental language use.

This immediately raises the question of the nature of 'conscious experience'. Thus, I begin here by setting out my assumptions about the relation between 'conscious experience' and brain activity. I do not offer a theory of consciousness. Rather, I focus on the nature of conscious behaviors. I take for granted that sense perception involves conscious experience, just as I take for granted that overt behavior involves muscular movement. My interest does not lie in how the brain produces consciousness or moves muscles, but in how bodily movement and conscious experience constitute unintentional behaviors, (intentional) actions and practices.

In the first section, I outline my assumptions, including an assumption of common cause: some brain activity which is sufficient for conscious experience is also necessary for the production of behavior. From the question of consciousness, I turn to the development of mental behaviors through the suppression of overt movement, paying particular attention to verbal thought. In the third section, I address the mental behavior of desiring — a mental behavior that is central to understanding intentional behavior, or
action. This will bring the discussion of the next chapter back to linguistic behavior, and in particular, intentional and deliberate speech acts.

Conscious Experience and Behavior

People are often conscious. For the most part, people are conscious exactly when they are capable of perceiving. Consciousness is a primitive: sensation (including internal sensation such as hunger, pain, etc.) just is conscious. That is how sensing happens in people, and probably in at least all vertebrates. I do not mean to say that we are conscious of everything we perceive or react to. People perceive things that they are not already aware of, and in fact, people even learn from things that they perceive unwittingly. But whenever we are conscious we are perceiving something and whenever we are perceiving anything we are conscious.¹

People perceive things when conscious. But not all of our consciousness seems to be perceptual awareness. For instance, we are often conscious of our thoughts, our feelings and our own behavior, etc. Although we perceive something in these cases, it does not appear to be the same kind of perception as occurs when we look at objects, hear sounds and feel hunger or pain. I give a more detailed discussion of this in sections two and three, but the upshot of my position is that any behavior whose production utilizes representational neuro-structures will thereby excite some of the same patterns of neural

¹It might appear that sleep and dreaming are exceptions to this. Dreaming, however, is a semi-conscious state in which one does perceive something, namely the elements of the dream. Sleep is more problematic. At least some monitoring of the environment must go on when one is asleep, since noises or cold can wake you up or cause you to move. This suggests that there are degrees of consciousness, and that sleepers are conscious, unlike people in a profound coma — the truly unconscious. On the other hand, although there is some monitoring of the environment, you don't really sense things when you are asleep. One only feels cold to the extent that one has woken up. Further complicating the matter is memory — it may be that one does feel cold when one is asleep but that one tends not to remember feeling cold when one wakes up. After all, most people accept that we don't remember most of our dreams, which implies that at least some conscious activity occurs when we are asleep that is forgotten by the time we awake.
activity that direct perception excites. After all, representational neuro-structures are caused by patterns of neural activity involved in perception, and the same patterns of activity (as caused the representational neuro-structures) occur in the utilization of those structures. Thus, mental behaviors such as thinking, feeling (emotions), imagining and deliberating involve conscious perception, albeit indirectly. Mental behaviors that we experience consciously are experienced consciously for precisely the same reason that we consciously experience overt behaviors — any behavior which utilizes representational neuro-structures involves sensation.

I take it for granted that the brain activity involved in ordinary perception is sufficient for conscious experience (again, this does not imply consciousness of all that is perceived or reacted to). Furthermore, I assume that the neuro-physiological activity sufficient for conscious experience is also necessary for some human behavior. Scientifically well-informed people have come to believe that there is a significant relation between the brain and behavior because countless scientists have correlated different activities, events or structures in the brain with different human behaviors. For instance, when certain parts of the brain are damaged, people can no longer behave in certain ways. When certain parts of the brain are electrically stimulated, people move in certain ways. Although it has certainly proved difficult for neuro-scientists to narrow down exactly what types of brain activity are necessary for exactly what types of behavior, there has been considerable progress. We believe, for instance, that in order for people to move there has to be some activity in the motor cortex, and that for people to talk, there has to be some activity in Broca's area.

There is also good empirical reason to assume that brain activity sufficient for conscious experience is necessary for certain kinds of behavior. The assumption is based
on subjective experience combined with beliefs (based on the empirical sciences) about the relation between brain and behavior. We have subjective evidence of our conscious experience, and empirical evidence of the correspondence between our having conscious experience and our behaving in certain kinds of ways. We are generally conscious when we are behaving. Exceptions may include sleepwalking and automatism, but even in these relatively rare cases it is not clear whether the behaviors themselves are not conscious (since people in such states may report that they are indeed conscious) or whether they are intrinsically non-recollectable. Furthermore, empirical sciences have given us good reason to believe that there is some sort of a connection between our brain activity and our conscious experience (which is why the question of the relation between mind and brain got raised in the first place). We are confident that some scientists can tell whether or not a person is conscious by monitoring events and conditions in the brain. They can do this because there is a difference between the kinds of things that happen in the brain when people are having conscious experience and the kinds of things that happen in the brain when people are not having conscious experiences. In short, the brain activities of people who seem conscious are different enough from the brain activities of people who seem unconscious for some scientists to tell the difference just by monitoring events and conditions in the brain. Indeed, we are confident that these scientists can determine whether or not someone is conscious by monitoring things in the brain even in the absence of behavior which we associate with our own conscious experience (for instance, when someone is paralyzed).

The kinds of behaviors that can occur when someone is not conscious (e.g. reflex behaviors) do not appear to involve the same kinds of brain events as those that occur when someone is conscious and engaging in non-reflex behaviors. Thus it appears that at
least some brain activity sufficient for conscious experience is also necessary for non-reflex behaviors. There is evidence that some brain activity is necessary for some human behavior. In addition, there is evidence that some brain activity is sufficient for conscious experience. Moreover, the same evidence suggests that some of the brain activity sufficient for conscious experience is also necessary for some types of behavior (non-reflex behaviors).

Yet something is wrong here. It is difficult to imagine any movement of an animal body (including human bodies) which cannot, at least in principle, be exactly reproduced through the use of external forces (including external forces such as the electrical stimulation of the brain, as we see in science fiction movies in which a mad scientist hooks up electrodes to a corpse). This suggests that there may be no type of brain activity that is necessary for any kind of movement. Furthermore, it is also evident that there is no type of brain activity sufficient for a kind of movement (since there can always be a break in the connection between the brain and the part of the body that is supposed to be moving). If brain activities are neither necessary nor sufficient for behavior, then what kind of connection could there be between brain activity and behavior?

Consider the following proposal which uses a simple mechanical analogy. The operation of the motor causes an axle to turn. Yet, the running of the motor is not sufficient for the turning of the axle since things can go wrong: there can be external forces acting on the axle, or the clutch can be disengaged, which prevents the motor from turning the axle even though the motor is running. Furthermore, the running of the motor is not necessary for the turning of the axle; for the axle can turn without the motor running, if there is another force acting it. The motor's turning is neither necessary nor sufficient for
the turning of the axle. What kind of relationship is there, then, between the running of the motor and the axle’s movement?

Although it is neither necessary nor sufficient that the motor be running for the axle to turn, it is necessary (but not sufficient) that the motor is running for the axle to be turned by the motor. We can distinguish between movements of the axle caused by external forces and movements caused by the motor. If I grab hold of the axe and turn it by hand, the movement of the axe is caused by a force external to the motor (me) and not by the motor. But if the motor is running, the clutch is engaged, and there seem to be no forces external to the motor acting on the axle sufficient to explain the turning of the axe, then we can generally assume that it is the motor that causes it to turn. Similarly, we can distinguish between movements of animals caused by external forces (mere movement) and movements of animals caused by forces internal to the animal (behavior).

In the motor-axle example, although the motor’s running is not sufficient for the turning of the axe, it is a trivial truth that the running of the motor is a necessary condition for the turning of the axe by the motor. It is a trivial truth because the distinction between the axe turning and the axe being turned by the motor cannot be made except by considering whether or not the motor is doing the turning. However, the distinction between animal behavior and mere movement can be made independently of considerations of brain activity, and thus, the fact that certain types of brain activity are necessary for certain behaviors will not be a trivial truth.

Behavior can be distinguished from mere movement in the following way: animal behaviors are overt movements initiated by the moving animal. ‘Initiated by’ is meant to capture the traditional external/internal distinction without invoking concepts such as responsibility, volition and intentionality, which often are associated with agency. I use it to
distinguish animal behavior (with an internal cause) from animal movements that occur when an animal is acted upon — running as opposed to being pushed along, kicking as opposed to having one's leg moved. The internal/external distinction is crude: internal means inside the skin and external means outside the skin. Theoretically, at least, one can distinguish behavior from other movements by treating the whole animal (meaning, the skin and everything in it) as a simple Newtonian body. To the extent that the animal's movements are accounted for by considering only external forces acting on the animal's body when treated as a whole, the animal's motion is not behavior, but mere movement. To the extent that the movements of the body cannot be accounted for merely by considering external forces acting on the body (as a whole), the movements can be considered animal behavior. Thus, when one considers the movements of a cat when it falls from a ledge, the movement of the cat downward can be explained as a result of the result of the force of gravity acting on the cat (as a whole), but the twisting of the cat (beyond that contributed by the initial spin) cannot be accounted for simply by considering external forces acting on the cat (as a whole); thus, it is animal behavior.

This definition of behavior also includes as behavior many of the things that go on in the body (for instance, digestion, blood flow, the pumping of the heart, etc.) since they cannot be explained simply by considering forces external to the body. Since many forces external to the body also act on things internal to the body, the same internal/external distinction must be used to distinguish mere heart movements (as might be caused by electrical stimulation of the heart by electrodes) from animal-initiated heart movements. Any movements internal to the body (such as contraction of the heart muscles) that cannot be accounted for simply by considering forces external to an animal, are behaviors of the animal, that is self-initiated movements.
Although this definition of behavior includes such behaviors as heart contractions and digestion, I will assume that these types of behavior can be distinguished from other types of behavior in that we have good empirical reason to think they can be demarcated from what we normally call behavior on the basis of the type of brain events which appear to cause them; for neuroscientists attribute these behaviors to a distinguishable system, the autonomic nervous system. Thus, from this point forward, I will focus only on the types of behavior that we traditionally have called behavior, self-initiated animal movements that are not taken to be caused by the autonomic nervous system.

Although it is quite possible that there are no brain activities necessary for any animal movement (since the same movements can be achieved through the use of external forces), it is quite plausible that certain kinds of brain activities are necessary (but not sufficient) for much animal behavior. Since the distinction drawn between behavior and movement in animals is independent of any consideration of brain activity, it is not a trivial analytic statement that certain brain activity is necessary for certain behaviors (as was the case above, when we said that the running of the motor was necessary for the turning of the axle by the motor).

I want to turn to the observation that an analogous argument can be made for the motor. If the motor is part of a car, then one can distinguish between the movements of the car initiated by the car (car behaviors) and movements caused by external forces (mere car movements) by treating the car as a Newtonian body, and considering the forces acting on it from the outside, as when it is towed away by a tow-truck. Thus, we are as justified in talking about car behavior as we are in talking about animal behavior. Surely the term ‘behavior’ has been watered down too much!
Yet, this objection is easily met for the car example. After all, the car doesn't do anything at all unless somebody is controlling it, either by driving it or it by controlling it remotely. Either the controller should be considered an external force acting on the car (in which case the car is moving, not behaving) or it should be considered internal to the car, and thus, part of the car or part of the car-controller complex. In the latter case, if the driver or remote controller is an animal, then there is no reason to deny that the car-animal complex is behaving since we accept that animals behave, in the sense that they initiate much of their own movement. Thus, in this case, behavior hasn't been watered down at all.

But what if the controller is non-animal? Excluding gods and aliens, machines are the only other kinds of things besides animals that are capable of driving cars. One can even suppose that the car has a complex computer within it and hence is a kind of robot, capable of behaving exactly as cars being driven by humans behave. Is it reasonable to say that such a machine behaves, in the sense of initiating its own movement? This depends on one's view about the role of the designer. If one considers the role of the designer as a sort of remote controller, remote in time as well as in distance, then one may consider the designer to be 'part' of the complex. Thus, the car-computer complex is moving (since without the designer it is not self-initiated), and the car-computer-designer complex is behaving. However, if one does not care about the role of the designer, one has no good reason to deny that the car-computer complex is behaving, since its movements are self-initiated.² In this case, given the complex movement of the car (which

²One could, of course, deny this and simply state that only animals behave, but this would be begging some important questions.
behaves just as car-human complexes behave), and the fact that it is self-initiated movement, 'behavior' hasn't been watered down at all. 3

Thus, it is reasonable to claim that some brain activity is necessary for some behavior. Furthermore, the brain activity necessary for some behavior is sufficient for conscious experience. The vast majority of observable human behavior, ordinary overt behavior, is precisely of this type: it requires brain activity sufficient for conscious perceptual experience. (Again, I do not claim that one is conscious of all that one perceives. Rather, when one is conscious, one is perceiving something, and conversely, whenever one is perceiving or behaving, one is conscious.)

Mental Behavior and the Suppression of Overt Movement

In cases of ordinary overt behavior, the patterns of brain activity involved in sensation (which is accompanied by conscious experience) trigger further patterns of brain activity (determined by representational neuro-structures) which produce muscular movement. Such patterns of brain activity which would normally result in overt movement can also fail to produce such movement. Things can go wrong. Consider a case of paralysis (due to spinal cord injury) in which the brain activity up to and including that involved in moving muscles occurs but does not produce muscle contractions. If this is a failed instance of overt, conscious behavior then some of the brain activity is sufficient for conscious experience, since ex hypothesi the failure did not concern any brain activity but

---

3 Of course, the role of the designer can also come into consideration with animals. If one considers God to be a designer, then whether or not anybody is really behaving depends on what one takes God's role to be and whether or not one thinks God can behave. Similarly, genes might be considered to be like designers, and genes are also initially 'external' to the body. Thus, this implies that animals including humans can behave (rather than move) only in so far as the genes are considered part of the body.
merely the muscle movement. Thus, despite the lack of overt behavior, these patterns of brain activity would result in conscious experience.

Similarly, overt muscular movement could be prevented from within. If, in addition to the patterns of brain activity that would normally result in muscle contractions, other patterns of brain activity interfered with, or suppressed, the muscle contractions, then although the brain activity that would normally result in overt behavior would occur, the associated overt behavior would not. Only when these secondary patterns of brain activity (which suppress the overt movement) ceased would the overt behavior take place, provided the primary patterns of brain activity were still occurring. Yet, during the period during which the overt movement was suppressed, all the brain activity which would normally result in overt movement would still occur. In such a case, the brain activity would be sufficient to produce conscious experience even though the associated overt behavior would not occur, since ex hypothesi, the brain activity is sufficient for conscious experience.

I argued in the previous section that many types of brain activity necessary for overt behavior are sufficient for experiential events. The brain activity involved in sensation is, for example, sufficient to produce conscious experience. We sense by being conscious. But, patterns of brain activity involved in sensation are the very patterns that cause and are activated by representational, perceptual neuro-structures (as discussed in chapter 3). Thus, the activation of such neuro-structures should also produce conscious experience. Furthermore, these perceptual neuro-representations form parts of complex neuro-structures which have as their objects the reconstructed objects of perception. As discussed earlier, the stimulation of a part of the complex neuro-representation will tend to cause stimulation of the entire complex. Since the complex partly consists of perceptual
representations (connected to one another), the stimulation the complex includes activation of perceptual neuro-structures which are sufficient to produce conscious experience. Thus, conscious experience should result whenever a perceptual representation is activated, whether or not the cause of the activation is sensory stimulation. Indeed, the performance of any ordinary, learnt behavior should result in some conscious experience of the behavior itself (and not merely of the present percepts), since learnt behaviors are produced through the activation of representational neuro-structures, which are first and foremost perceptual neuro-structures and thus sufficient for conscious experience.

For instance, a person’s perceptual neuro-representations of dogs (and of dog-experiences or observed dog-related events) may be incorporated into complex neuro-representations which take as their object the reconstructed objects of experience which caused these representational structures, namely dogs and dog-related events. Learnt behaviors, such as petting a dog, may draw on these complex neuro-representational structures. The patterns of brain activity involved in the behavior of petting the dog could therefore include activation of many of the perceptual neuro-structures which comprise the complex representation, which will also be sufficient for conscious experience. Petting the dog may therefore cause one to have a conscious experience of one’s behavior ‘petting dog’, since it activates a representational structure formed by many observed dogs, and dog-petting events. Thus, in addition to having conscious experience of the dog and the surroundings when petting the dog, one may have a conscious experience of one’s own behavior — of what one is doing. The representational structures utilized in the production of behavior are perceptual representations (at least in part) and hence are sufficient for conscious experience.
This conscious experience is not necessarily linguistic or even quasi-linguistic. However, as discussed in the previous chapter, some representations of word sounds are fully integrated into complex representations of the reconstructed objects of experience. These word-object representations may also be connected to other complex neuro-representations which could be utilized in the production of behaviors such as petting a dog. Again, if these behaviors result in the activation of the entire representational complex then the behavior of petting the dog could cause the activation of the word sound representation, which is sufficient for a conscious experience akin to the auditory perception of the word sound 'dog' or 'nice dog'. This conscious experience may not be as vivid as that caused by hearing the word 'dog' spoken, since some perceptual activity which may normally be caused by hearing words spoken may not occur. In addition, the indirect trigger (by the activation of connected representations) may not produce as much activation as a direct trigger (hearing the word spoken).

Although many stable neuro-representations are caused through repeated exposure to like objects and events, there is no reason why relatively stable neuro-structures or connections between neuro-structures should not be formed from a single exposure. Indeed, patterns of connections caused by recent incidents or incidents that cause unusual patterns of activity would not have had much of a chance to be destroyed or altered by subsequent brain activity. Thus, there can be relatively stable representations of single episodes or events, and relatively stable connections with other neuro-representations even after a single association. Thus, the patterns can be activated by perceptual stimulation, or by the performance of behaviors which utilize representational structures which are connected to them. In other words, seeing things or engaging in learnt behaviors may activate related perceptual representations which will produce conscious experience similar to that experienced in the past — it may cause conscious
experience of remembered episodes. And as discussed in chapter 3, whenever a relatively unstable representation is so activated, it helps to stabilize it. Not only can one remember single episodes, but the remembered form of the episode becomes more stable the more it is recalled or utilized.

An ordinary overt behavior can produce conscious experience in addition to the sensations that occur during the performance of a conscious behavior — for instance, conscious perception of what one is doing, consciousness of associated words and consciousness of remembered episodes or perceptions. Thus, if the muscular movements involved in overt behavior were suppressed by further brain activity, these conscious experiences associated with the production of the behavior (rather than the changing perceptual input that might be caused by the behavior) would remain. Therefore, even mental behaviors (internalized overt behaviors) would produce conscious perceptual experience independent of the conscious experience caused by present perception. In other words, when one thinks in words or visualizes an object one may have a conscious perceptual experience similar to hearing words or seeing objects, since the behavior would utilize, and thus cause activity in, perceptual neuro-representations. This conscious experience would be independent of any further conscious experience of sensing one's environment while performing the mental behavior. I can think in words, feel hungry and hear the sound of the computer's motor humming all at the same time.

Whether or not this account is neuro-physiologically accurate, it shows that the occurrence of suppression of overt movement and its potential for transforming overt behaviors into mental behaviors through internalization can be readily accounted for on a connectionist, representational model of the brain. Given that perception is conscious, no other form of consciousness needs to be considered. The phenomenal aspect of mental
behavior can be fully accounted for in the same way that the phenomenal aspect of overt behavior can be — behaviors utilize perceptual representations, and thus much of the same kind of perceptual brain activity that might occur during sensation (thereby causing these perceptual neuro-representations) would also occur during the later performance of mental or overt behaviors.

To summarize the notion of suppression, overt movements are suppressed when the brain activity involved in the overt behavior (which consists of those overt movements) persists but additional brain activity interferes with the performance of the overt, muscular movements. The resulting brain activity produces a mental behavior. This mental behavior is much same as the related overt behavior, except that in the case of the mental behavior the overt movements are suppressed. Thus, the conscious experience involved in the overt behavior is similar to that involved in the mental behavior in so far as much of the brain activity sufficient to cause conscious experience is shared between the mental and the overt behaviors. Thus mental behavior can be understood as an internalized overt behavior.

Internalized Speech

There is little doubt that some mental behaviors that involve verbal thought, such as silent reading or silent prayer, are internalized overt behaviors. A child learns to read and to pray aloud. Yet, at some point, a child might be expected to perform such an activity 'silently'. Many children who learn to suppress the overt movements associated with these activities go through a stage of incomplete suppression. Children learning to read silently may first read in whispers, then mouth the words before finally succeeding in silent reading. Moving one's lips while reading is in fact a mark of an inexperienced reader.
my own experience, when class recitation of the Lord’s Prayer was eliminated from opening exercises and children were given the chance to engage in silent prayer, many children whispered or mouthed prayers. Likewise, if one asks a child to think of song and just clap its rhythm, he will often try to sing the song ‘silently’. Many young children fail to completely suppress the words in this task, and either mouth or whisper them.

When we first perform mental behaviors that are clearly modeled on overt behaviors such as speaking, reading, praying or singing aloud, we often go through a stage of incomplete suppression of the overt behavior — whispering or moving our lips. The fact that we go through this stage suggests that, at least in these activities, the mental behavior is the counterpart of overt behavior with the final stage of overt movement suppressed. Not all children go through the stages of incomplete suppression — some may just be better at suppressing the overt behavior.

If specialized cases of thinking in words (silent reading, prayer or song) are internalizations of overt verbal behavior, then this suggests that thinking in words itself is suppressed overt speech behavior. Certainly many young children seem to go through stages in which they talk to themselves, whisper to themselves or mouth words. These too, may be stages of incomplete suppression, normal (but not necessary) stages in developing the ability to think in words. There is little doubt that thinking in words and speaking aloud draw on the same resources. Indeed, it is quite difficult to think in words while speaking different words. If one rehearses a speech, memorizing it so completely that one can speak it without thinking about it, then one can think something else while speaking. One can also make rapid switches between speaking aloud and thinking, speaking a sentence or phrase, quickly thinking something else, and then continuing one’s sentence. Indeed, we often interrupt our overt speech with thoughts. But, for the most
part, we cannot think one thing while simultaneously speaking something else. Thinking in words and talking seem to draw on the same resources. Thinking in words appears to be internalized speech.

Furthermore, other phenomena suggest that thinking in words involves active suppression of overt behavior. By active suppression, I mean that all the brain activities involved in the overt behavior occur, and additional brain activities suppress or interfere with the overt, muscular movement. If suppression is active, then since the activity involved in producing the behavior is all occurring, the conscious experience of the behavior would be complete. That much suppression is active in this way is suggested by the fact that 'slips' in suppression result in the activity which was meant to be suppressed. If suppression is active, then the brain activity producing the behavior is ongoing. Thus, if there is a slip in the suppression, the suppressed activity continues in its overt form.

Sometimes we think to ourselves, we slip and speak aloud what we are thinking, sometimes without even being aware that we are speaking aloud until someone draws it to our attention. This generally occurs when we are excited, afraid or otherwise agitated about something (perhaps, excited by the idea that occurs to us or afraid that something bad will happen) or when we are extremely relaxed (or falling asleep). Both of these states, agitation and extreme relaxation, are associated with lack of control. When there is too much or too little going on in our minds, we 'forget' to suppress the overt behavior, resulting in our saying aloud what we were thinking. Thus, if lack of suppression can be caused by 'overload' it also explains why people often mouth or whisper when they are concentrating hard on something new, for instance, reading a difficult passage or doing a difficult math problem. Excessive talking, whispering, muttering and mouthing words to oneself (rather than keeping one's thoughts silent) are often associated with mental illness,
senility, drunkenness and young children — all of which are also associated with lack of control. When we lose control, we speak what we would otherwise think. This suggests that thinking involves active suppression of overt behavior — when we have not yet developed control, or when we lose control, we don’t successfully suppress the overt behavior.

The notion of active suppression also serves to distinguish this view of verbal thought from that of the behaviorists such as B. F. Skinner and John Watson. Like the behaviorists, I am claiming that verbal thought is some form of internalized speech. Yet, behaviorists held that verbal thought was in some way a weaker version of overt speech. Watson believed that when we think in words we produce weak muscle contractions in the larynx, mouth, tongue, throat, chest, etc. He writes, “My theory does hold that the muscular habits learned in overt speech are responsible for implicit or internal speech (thought).” (Watson 1930:239) Skinner described verbal thought as covert linguistic behavior; a behavior becomes covert on his view when “its strength drops below the value needed for overt emission.” (Skinner 1957:435) In contrast, I am not claiming that verbal thought is weakened overt speech; rather, verbal thought requires something in addition to ordinary overt speech — the active suppression of the muscular movements involved in overt speech. In addition, the active suppression of overt movement is not limited to speech behaviors, but is an important feature in the development of all mental behaviors (verbal and non-verbal alike).

Desiring

Partial suppression is also evident in other behaviors. Consider a child who has learnt to associate the command ‘No!’ with stopping whatever she is doing. If, for instance,
she is told 'No!' when reaching for something dangerous (the fire in the fireplace), she stops. Yet, although she stops, she does not seem to stop completely — she partially suppresses her reaching behavior. While falling short of reaching for the flame, her behavior remains oriented to it. Her body is turned towards it and her arm muscles are tensed as if she is about to reach for it again. Indeed, after a bit of time passes she reaches out for the flame once more. Again, at the command "no!" she retracts her hand. This continues until the speaker loses patience with this "game" and picks the child up.

Most children, at some point, seem to play these "games", although the command need not be "No!" but could be some other spoken or physical indication that the children have learned to associate with stopping their object-oriented behavior. The fact that the object-oriented movements are only partially suppressed suggests that whatever is going on in their brains that normally results in the object-oriented movement continues even when the brain activity that is triggered by the command "No!" interferes, and prevents the completion of the behavior, the muscular movement. Young children can be observed partially suppressing their object-oriented movements in a wide range of circumstances. Many children, when seeing a previously forbidden object, will suppress their own movements, at least for a few moments, even in the absence of a command. Likewise, when spying an object that attracts them but that in the past was associated with pain or fear, children will often hesitate before reaching or stop in the middle of reaching.

One important result of the suppression of object-oriented movements is that a time lag is created during which more information can be gained. Consider a child who spies something which she normally would grab. The suppression of the object-oriented movement temporarily stops the child from grabbing the item. In the period of time in which the movements are actively suppressed (the brain events that normally result in it are
occuring but the overt movements are not) she has an opportunity to receive more information about the world. The child can be said to be in suspense — some brain activity is causing grabbing behavior and other brain activity is causing suppression of the grabbing movements. Any new perceptual information can end the suspense by pushing her one way or the other — to end the brain activity involved in suppression (thereby causing her to grab the object) or the end the brain activity involved in grabbing the object (thereby causing her to cease being oriented towards that object). The period of time in which she can gain new information increases the chances that what she does is beneficial for her. Whether she follows the suppression of the object-oriented movement with the overt object-oriented behavior or not, the suppression is to her advantage.

The suppression of overt object-oriented movements should produce a conscious experience similar to the conscious experience involved in performing the overt, object-oriented behavior. In fact, the conscious experience should be quite similar to the feeling of desire. Desire is object-oriented, whether or not there is any movement towards the object. Desire is often (but not always) followed by object-oriented behavior. Desire can be experienced when one is engaging in overt object-oriented behavior as well as when one suppresses those movements. And, most importantly, we often can tell what people desire from their body language and facial expressions. They lean a little towards what they desire, or look at it in a certain way, they sometimes tense their muscles as if they are about to move towards it, etc. When people behave as if they are partially suppressing object-oriented behavior, we assume that they desire the object to which their behavior is oriented.

Finally, the ability to suppress one's behavior, to resist acting on a desire, is a mark of maturity, responsibility or of having control over oneself. In the discussion of suppressed language behavior, thinking in words and silent reading, it was noted that as
people mature, they are generally able to suppress the behavior more fully. When young, or in situations in which there is a loss of control (excitement, extreme relaxation, insanity), the behavior is only partially suppressed, resulting in observable behavior such as the moving of the lips or whispering. Occasionally there is a complete failure, resulting in talking aloud when the person meant to be thinking or reading silently. This is also the case for the suppression of overt object-oriented movements, such as reaching for things or going after things. Young children are expected to reach for or grab anything that catches their eye, and as they mature, they are expected to suppress these movements. The younger a child is, the less fully the movements are suppressed and the easier it is to tell what he or she wants from his or her facial expression or body language. An adult suppresses his object-oriented movements more completely. Yet, when an adult loses control, due to excitement, insanity or inebriation, it is often easier to see what he wants. In such situations, the object-oriented behavior is less completely suppressed, and thus the facial expressions and "body language" are more revealing. And like a systematic failure to suppress one's overt speech, a systematic failure to suppress one's object-oriented movements (acting immediately every time something catches the eye) is a mark of the young, the senile, the insane, the inebriated and others who do not have control over their behavior.

It seems the internalized object-oriented behavior which people evidently engage in is a form of desire. But 'desire' is ambiguous between the conscious experience that accompanies any object-oriented behavior (overt or internalized) and the internalized object-oriented behavior itself — the mental behavior of desiring. The first type of desire (simple desire) is experienced by any animal that perceives consciously and utilizes representational neuro-structures in the production of object-oriented behavior, whether or
not the animal is capable of suppressing its object-oriented movements. Presumably, neonates and all vertebrates experience simple desire when they perform conscious, object-oriented behaviors. Simple desire does not involve any higher-order awareness of the desire — one need not have a concept of desire or know that one feels desire in order to feel it. Desire, in this sense, is just a name for the experience that a conscious creature has when it engages in object-oriented behavior.

The more interesting forms of desire, however, require the suppression of object-oriented behavior. The internalized object-oriented behavior caused by the suppression of overt object-oriented movements can be referred to as the mental behavior of desiring. Like overt object-oriented behaviors, the mental behavior of desiring need not involve any higher-order awareness of the desire. Often, however, we are very clearly aware of our desiring — we have a higher-order awareness of it. Suppression, which can produce the mental behavior of desiring, is necessary for higher-order awareness of mental behaviors such as desiring. Overt behaviors that are experienced consciously have a mental component that is shared between overt behaviors and their mental counterparts. The conscious experience will include the feeling of simple desire, and also an awareness of the behavior engaged in (as discussed in section 2). But unless the behaver has the capacity to suppress overt movements then all conscious experience of the desire and of the behavior is co-extensive with the overt behavior. Desire and awareness of the behavior, in this case, would always co-occur with each other and with the behavior itself. Thus, no distinction can be made between the desire and awareness of the behavior, or between the behavior itself and any conscious experience. Without the ability to make these distinctions, one cannot reflect on the desire or on the awareness of the behavior. Therefore, suppression of certain overt movements (the mental behavior of desiring) is necessary for higher-order awareness of conscious experiences associated with behavior.
including desiring itself and awareness of the behavior, since it allows one to make such a distinction.

When a conscious individual suppresses object-oriented movements, the brain activity of the internalized behavior is sufficient for a feeling of desire as well as an awareness of the behavior that is being suppressed. Consider, for instance, a child who sees a book which triggers object-oriented behavior as well as triggering suppression of that behavior (due to past association with the command 'No!') She engages in the mental behavior of desiring the book. In doing so, the child has an awareness of the suppressed movements, reaching for the book, since the neuro-representations utilized in reaching for the book are activated — only the overt movements are suppressed. In addition she will have a sense of simple desire. (However, she need not have a higher-order awareness of either her desire or her awareness of the behavior.) The child, in suppressing the movement, hesitates before reaching for the book. However, during her period of hesitation she has more perceptual experiences. These experiences (perhaps seeing her mother looking the other way or discovering that no command "No!" is forthcoming) cause the suppression activity to cease and the underlying book-oriented behavior continues without interference — she reaches for the book.

Her behavior in reaching for the book is caused by same brain activity that occurred during the period of suppression. During the overt movement, however, no further brain activity interfered with the movement, whereas during the period of suppression, further brain activity did so interfere. It is significant that the reaching movements were caused by the same brain activity that constituted the mental behavior of desiring, since this brain activity was also sufficient for a feeling of simple desire and an awareness of the behavior prior to reaching. In other words, the child reaches for the book because of brain activity sufficient for a feeling of desire and for an awareness of the behavior that would fulfill the
desire (reaching for it). This is quite close to the traditional conception of action, or intentional behavior — a behavior is intentional if it is caused by a desire or a pro-attitude and by a belief that such a behavior would obtain the desired objective.\(^4\)

One advantage of considering the role of suppression in transforming an overt behavior into a mental behavior is that there is a neat neuro-physiological distinction that can be made between unintentional behaviors and intentional behaviors or actions. Although there are still a number of details that need to be discussed, in broad outline, those behaviors which follow a period of suppression are intentional. These actions are caused by brain activity sufficient for desire and awareness of the behavior that would fulfill the desire. In order for this approach to succeed, suppression must itself be unintentional. If suppression is itself intentional then a regress of suppression results — intentional behaviors are previously suppressed behaviors, but suppressed behaviors are intentional behaviors which are previously suppressed behaviors etc. Therefore, it must be clearly established that suppression is not itself an intentional behavior.

Consider the following example of suppression and the subsequent behavior:

A toddler spies a book within her reach. The child begins to reach for it, but then hesitates. In the past her parents have always said 'No!' and taken the book away whenever she reached for one, and she has come to associate reaching for books with something negative. But, nothing happens during the moment of hesitation. She looks at her parents as if to check whether or not they will say anything. They don’t. Then she reaches for the book, thereby performing an action.

I want to show that it is possible for the child to perform the act without supposing that she performs any other action, such as an act of suppressing, or an act of hesitating or deciding. What is necessary for the child to behave in this way?

\(^4\) For my purposes, the tradition can be thought of as starting with Aristotle and represented recently by Anscombe and Davidson.
To begin with, the child must have a neuro-representation of the book perception which is itself a part of a more complex neuro-representation of reaching-for-books. When the child sees a book, this then triggers the activation of the entire complex which would normally result in the child reaching for the book. However, in the past when the child reached for a book her parents commanded 'No!'. Thus connections formed between the neuro-representation of the complex reaching for books representation and the neuro-representation of the word sound 'No!'. Furthermore, the representation of the word sound, 'No!', is part of a complex neuro-representation of stopping behavior. On numerous past occasions the word has co-occurred with stopping behavior — the child's parents frequently said no while simultaneously stopping the child. Thus, the sound 'No!' causes activation of the neuro-representation of stopping behavior. Thus, the seeing of a book triggers the book representation as well as the reaching for the book representation which causes the child to begin reaching for the book. But the activation of the reaching-for-book representation causes the activation of the connected 'No!' representation which is itself part of the complex stopping representation which is thereby activated. Thus, two important neuro-representations are triggered at once — the reaching representation utilized in reaching behavior and the stopping representation utilized in stopping behavior. The latter functions to suppress the overt movement of reaching for the book. However, the reaching representation is still activated. In fact, the child only partially suppresses the reaching movements, remaining oriented towards the book.

During the period of time when both the stopping and the reaching neuro-structures are stimulated, the child is in suspense. This creates a gap during which new perceptual information can have an effect. The 'No!' representation is also connected to other neuro-representations, like representations of her parents. These further representations may be utilized in other behaviors, and thus their activation might cause the child engage in related
behaviors, for instance, turning to look at her parents. The observation of her parents strengthens the parent representations. Although the parent representations are connected to the ‘No!’ representation, in this instance, there is no occurrence of ‘No’. Thus causes a relative weakening of the ‘No!’ representation, which results in a weakening of the stopping representation to which it is connected. This means that the stopping representation is no longer being utilized in the suppression behavior. Thus, the overt movement is no longer being suppressed, and the final stage of the reaching behavior is realized — the child reaches for the book.

The suppression process does not involve any intentional behavior, or even conscious awareness on the part of the child. Indeed, this type of suppression can easily be induced in animals such as dogs. One can train a dog to fetch balls when they are thrown. One can also train the dog to stay on the command, ‘Stay’, and not go until given a release signal. If one systematically says ‘Stay’ while simultaneously throwing the ball, the dog suppresses the fetching behavior (but only partially, as it will stay oriented towards the ball). Then, when the dog receives the further release signal, the suppression behavior ceases and the dog runs after the ball. If the release signal is not given, eventually some other percept will cause either the suppression or the ball-oriented activity to cease. Once trained in this way, the dog may suppress the fetching behavior when one throws the ball even in the absence of the command ‘Stay’. The fact that one can induce suppression indicates that intention is not required for its occurrence. But suppression also occurs naturally in a wide variety of circumstances, especially when a behavior frequently co-occurs with pain or unpleasantness. Indeed, by adulthood suppression of most object-oriented behavior is automatic — we don’t just go for things the moment we see them. It is
the rare instance when we have to intentionally stop ourselves from acting immediately on what we see.5

Since suppression can be fully accounted for without intentionality or indeed, any mental phenomenon, it can function without circularity or regress as the distinguishing feature of an intentional behavior. As discussed earlier, when the overt movements of a conscious animal are suppressed, the resulting internalized behavior is a mental behavior. The brain activity involved in the mental behavior is sufficient to produce a conscious experience akin to that associated with the overt behavior. When overt object-oriented movements are suppressed, the result is the mental behavior of desire. The mental behavior of desiring involves a feeling of simple desire and an awareness of the behavior which is being suppressed and would fulfill the desire. If the suppression ceases, the resulting overt behavior is an intentional behavior. It is caused by the very brain activity that was sufficient for the desire and the awareness of the behavior would fulfill the desire. Simple intentional behavior of this type need not involve any higher-order awareness of either the desire or the behavior. However, as discussed earlier, it also makes such a higher-order awareness of the conscious experiences possible.

Suppression of object-oriented behavior also creates the opportunity for choice. A child whose object-oriented movements are not suppressed will reach immediately on seeing something that attracts her, thereby triggering the reaching behavior. The behavior is automatic — the activation of the neuro-representations will result in the overt movement, provided that nothing interferes. Suppression provides interference. When

---

5 Most suppression of object-oriented movement is habitual in most older children and adults. I will deal with this in more detail in the next chapter when discussing habit formation and its relation to intentional and non-intentional behaviors. I contend that the mental behavior of desiring is primarily due to the habitualization of the non-intentional suppression of overt object-oriented movements. But, some suppression of movements is intentional. For my purposes here, it matters only that some suppression (especially that in young children) can occur without intent, either naturally (due to past association with pain) or induced (saying ‘no’ or slapping a child’s hand).
there are two competing patterns of neural activation, one causing the reaching behavior and the other causing the suppression of the movement, an opportunity for selection is created. The person must eventually do one thing or the other — go for it (stop suppressing and resume the movement) or to do something else (stop the mental/internalized behavior). The trigger object did not just cause object-oriented behavior, but it simultaneously caused suppression behavior. The trigger object caused a dilemma — the person must act or must cease to be oriented to that object. At the point of suppression, two options become available.

In contrast, an organism that engages in object-oriented behavior every time it encounters an object that triggers certain brain events cannot make choices. Such an organism goes for the first attractive thing it senses, whether or not there is something bigger and better nearby. It acts automatically whenever it comes across something that triggers object-oriented behavior. The organism that can suppress overt movement when encountering an object that would normally trigger object-oriented behavior creates an opportunity for choice. There is a possibility that something bigger and better will be sensed and acted upon instead, since there is time for more information to be received. There is a possibility that new perceptions of the trigger object will trigger an end to the suppression or an end to the object-oriented behavior itself — further perceptions of the object will result in engaging in the overt movement or stopping altogether. Since the mental behavior of desiring is the experience of suppression and since the ability to suppress overt object-oriented movements is necessary for choice, the mental behavior of desire is necessary for choice for conscious individuals.

---

6 Such an animal would have some discrimination in that they would only 'go for' some types of things. Frogs, for instance, must have a neuro-representation utilized in bug-snatching movements that are triggered (and thus represent) by flylike specks. Thus, they go for flylike specks but not diamonds or sunflower seeds.
The most primitive of intentional behaviors, the simplest actions, are the movements that follow from suppressed object-oriented behavior. In such cases, the actor does not intentionally suppress his overt movements. The suppression just occurs, likely because stopping behavior was associated in the past with the object-oriented behavior and, thus, the neuro-representations utilized in the production of both behaviors are connected and triggered together. The suppression of the overt movement results (in conscious individuals) in the mental behavior of desiring. Desiring serves three important functions: first, it creates an opportunity for selection by producing in the actor two competing brain activities which delays overt movement until further events determine which of the two brain activities will cease; second, it allows the feeling of simple desire and the awareness of the behavior to precede the overt movements; and third, by producing a feeling of simple desire and an awareness of the behavior distinct from the experience of engaging in the behavior, it makes possible a higher-order awareness of these conscious experiences.
Does it Act?

The most fundamental of intentional behaviors, rudimentary actions, are the overt behaviors which continue after a period during which the overt, object-oriented movements are suppressed. In such cases, the actor does not intentionally suppress his overt movements. The suppression occurs because in the past the object-oriented behavior was associated with stopping behavior; thus, the same object triggers both behaviors (or the object triggers object-oriented movement which then triggers the stopping behavior). The suppression of the overt, object-oriented movement results (in conscious individuals) in the mental behavior of desiring. Desiring serves three important functions: first, it creates an opportunity for selection by producing in the actor two competing behaviors (one which produces movement, and one which suppresses it). This prevents overt movement until further events determine which of the two behaviors will cease; second, it allows the feeling of simple desire and the awareness of the behavior to precede the overt movements; and third, by producing a feeling of simple desire and an awareness of the behavior distinct from the experience of engaging in the behavior, it makes possible a higher-order awareness of these conscious experiences.

The archetypal rudimentary action looks something like this: a child sees something attractive and starts to respond, stops short of reaching for it because in the past it was forbidden, looks around, sees no one to stop her, then goes for it. This picture of action is in keeping with the truism that a person acts (rather than responds
automatically, or moves involuntarily) if the person had the capacity not to act. Hesitation, or the few moments between starting to respond and completing the movement, provides evidence that the person had the capacity not to act (although it does not establish it).

A rudimentary action is best contrasted with automatic and unsuppressed reactions, like the smile that spontaneously lights up a person's face when she sees a friend. Many behaviors just seem to occur in certain circumstances: we recoil when we see something horrid, flinch when we perceive something coming our way, become sexually aroused with the right kind of perceptual stimulus\(^1\), freeze when afraid, laugh when amused, relax when a threat abates. In such cases, a perception almost mechanically triggers the behavior, sometimes even before we are aware of what we perceive. The behavior of young children and animals often appears to be of this type — babies just reach out at some of the things that catch their eye, frogs just zap their tongues at flying specks, children just grab a toy that someone else has taken, and puppies just chase certain things that move.

Some automatic reactions, such as recoiling, flinching, sexual arousal, smiling and staring, are not as obviously object-oriented as reaching or chasing. These behaviors may merely involve change in bodily position, or even involve movement away from the objects that they are oriented towards. Yet, the behaviors are still oriented towards their triggers. At the risk of sounding paradoxical, my recoiling is oriented towards the thing from which I withdraw when recoiling.

Negative object-oriented behaviors can stop other behaviors, for instance, recoiling in horror can stop an advance forward. Thus, such behaviors might seem to be suppression behaviors, such as stopping oneself from reaching or freezing in the middle of

---

\(^1\)For everyone, it seems, there are perceptual stimuli that cause arousal. But, people are aroused by different stimuli, just as people recoil from different things. Yet, there are cultural norms — North American heterosexual women are portrayed (at least in pulp fiction and on television) as 'melting' when they see attractive men looking at them in a certain way.
reaching. But, suppression behavior involves suppressing muscular movement oriented towards that object — stopping when one would otherwise be reaching for it, keeping silent when one would otherwise be asking for something. When one recoils at the sight of something horrid, one would not otherwise move towards the horrid thing. Suppose, for example, one is reaching in the back of the fridge for a jar of pickles and one touches something rotten causing one to recoil. Although I am moving in the direction of the rotten cucumber, my behavior is oriented towards the jar of pickles that is near it. I am in no way attracted to the rotten vegetable. Rather, touching it causes an automatic negative response to it — I recoil. Thus, my automatic recoiling from an object can be thought of as oriented towards the object, albeit negatively.

These negative object-oriented movements can in fact be suppressed, just as we suppress our positive object-oriented movements such as reaching. We can learn not to recoil at the sight of an insect or snake, not to tense up when a disliked person approaches, not to become sexually aroused in response to certain situations, and even not to blush when embarrassed. During suppression, only the overt movements are suppressed, but the behavior continues. Thus, the conscious experience during suppression is akin to recoiling (like a feeling of disgust) and includes an awareness of the behavior whose movements are being suppressed. I can force myself not to tense up in certain situations, but I still feel something negative oriented towards the person whose presence I dislike and I feel like leaving the person's presence. I can force myself to act normally when I touch rotten vegetables, but I still feel disgust and feel like recoiling when I do so.² It is a curious accident of the English language that the idiom 'feel like doing such and such' means 'am inclined to do such and such' rather than the literal (as I mean it) 'is

² Of course, we also train ourselves not to react at all to different stimuli or situations, which is different from learning to suppress our reactions to those situations.
similar to the feeling of doing such and such'. My point is that the feeling one has when one is inclined to do such and such is similar (in some ways) to some of the feeling one has when doing that which one is inclined to do.

The implication of our being able to suppress these object-oriented behaviors is that the movements which follow an interval of suppression should also constitute rudimentary actions. Thus, if I can suppress my immediate withdrawal at the sight of a disliked person, then what I do next (retreat or continue) is intentional. When I am suppressing withdrawal, recoiling, fleeing or flinching behaviors, two relevant behaviors occur — the negative object-oriented behavior and the suppression of the overt movements. The interval, once again, provides an opportunity for new perceptual information. If I perceive that it is harmless, then I may cease the recoiling behavior, but if I perceive that it is indeed horrid, then I may cease the suppression behavior and allow the withdrawing movements to continue. At the point at which I begin to suppress the recoiling behavior, two real possibilities for future behavior open up — I have to do one or the other eventually, withdraw or not. In addition, during the suppression interval, since the recoiling behavior is still occurring, I should have a negative object-oriented conscious experience as well as an awareness of the withdrawing movements which I am suppressing. If I then withdraw, the withdrawal follows from the feeling of disgust (or some other negative object-oriented experience) and an awareness of withdrawing (a perceptual belief about the behavior which satisfies the feeling of disgust).

A rudimentary action, then, consists of that overt behavior which follows a period during which the object-oriented movements were suppressed. 'Object-oriented' can be

---

3 Of course, other behaviors may also be occurring, for instance, reaching for pickles, walking to the store, etc.
4 Since both behaviors are occurring simultaneously, I do have to stop one or the other eventually, if only because I die. If I stop the suppression, I withdraw overtly. If I stop the recoiling behavior itself, I stop withdrawing altogether.
interpreted literally: reaching, eating, chasing, fleeing, smiling, getting aroused, flinching, withdrawing, fighting, hitting, staring, et cetera can all be oriented towards objects. Some overt movements may be more difficult to suppress (getting aroused, flinching) and others stand out as behaviors which we are trained to suppress when we are quite young, for instance, reaching, grabbing, hitting and staring. Suppressed object-oriented behavior will tend to give rise, in conscious individuals, to feelings or attitudes that correspond to the suppressed behaviors. Other people make assessments of what we feel on the basis of our partial suppression as well as from unsuppressed overt movements, and in this way we learn names for these feelings. Thus, suppressed hitting, fighting and kicking is usually considered anger, suppressed staring or touching is interest or curiosity, suppressed sexual behavior is lust. These may be felt both during the suppression interval and during the unsuppressed behavior. But, it is because we can feel them in isolation from the overt behavior with which we associate them that we can identify them as feelings independent from those behaviors.

Not all actions are rudimentary. I call these rudimentary actions because they are the first actions from which more sophisticated actions develop. In chapter five, I distinguished between movement and behavior. Here, I distinguish between mere behavior and action (whether rudimentary or sophisticated). An obvious implication of this is that a clean distinction can be made in theory, if not in practice, between creatures that can act and those that merely behave. (However, a clean distinction cannot be made between actions and mere behaviors.)

In order to act one must be able to engage in at least two behaviors simultaneously — one behavior that would normally produce overt, object-oriented movement and one which suppresses the overt movement. The behaviors must be self initiated (as discussed
in chapter five) and triggered by sensory stimulation (or the analogue of that for machines).

In order for suppression to occur, the creature must be able to learn — it must utilize representations in the production of behavior. Otherwise, it would continue to respond to perceived objects in accordance with its initial dispositions. Suppression is a new way of responding to a previously encountered type of stimulus, and suppression utilizes representations that were created when the creature previously encountered stimuli of that type.

I am making a strong distinction between actors and non-actors. Yet my own interest is not in determining which creatures are which, something which can only be done empirically. I do think that we have good reason to hold that all mammals can act. They can learn and, in fact, they exhibit what looks like partial suppression behavior. And perhaps all vertebrates can act. Invertebrates, even bees and ants, probably don't act, but again, this would be an empirical question. Many invertebrates certainly learn. The bee must utilize representations in the production of her figure-eight dance. Perhaps people who study bees will find evidence which suggests that bees can learn to suppress their dance behavior in some situations, and only go on to perform the dance in response to further perceptual information. Likewise, although I doubt if any machines act, there is no reason, in principle, why a machine should not be able to. Suppression, however, is something to look for when asking the question, “Does it act?”

But not all actions are rudimentary. The capacity to perform other actions, however, develops from the capacity to perform rudimentary ones. These other actions, including speech acts, mental acts, deliberate or planned acts and habitual acts (and all combinations, such as planned speech acts and habitual mental acts) do not necessarily consist of movements that follow an interval of suppression. Yet, suppression is central to
understanding the development of actors: development from performers of rudimentary actions to speakers, deliberators, promisers and baseball players.

The remaining part of this chapter addresses the development of actors with respect to habit formation, speech acts and deliberate action.

Habit Formation

Churchland (1986: 151-152, quoted above in chapter 3) advocates the use of the term ‘plasticity’ when discussing various forms of learning.

Among kinds of nervous system plasticity already believed to be distinct phenomena are habituation, sensitization, classical conditioning, operant conditioning, imprinting, habit formation, ...

Her list extends to include most forms of what we would call learning. In this essay many of these forms of plasticity are made the subject of philosophical analysis. But I have little to say about habit formation other than, “It happens.” Children learn skills — reaching, focusing their eyes, rolling over, holding up their heads, sitting up, walking, etc. They learn to make phonemes, words and sentences. They learn to run, drink from cups, turn on the tap, tie their shoes, ride bicycles, read and write, brush their teeth.

I take it for granted that this kind of skill acquisition or ‘know-how’ is habit formation. The more you do something, the better you get at it (at least up to a point). I take it for granted, too, that the reason that one gets better the more one does something is that the patterns of brain activity involved in the habitual behavior stabilize. The process is much the same as that outlined earlier in the discussion of the formation of neural representations. The more that overlapping patterns of neural activity are produced, the more likely that those patterns will be produced in the future when they are triggered by associated brain activity.
There are a few important points I do want recognized about habit formation. First, when one performs a behavior habitually, one may stop being aware of the habitual aspects of what one is doing. For instance, when I am riding my bike, I am generally not aware of the habitual aspects of my bicycle riding. I am aware of where I am going (unless I go there out of habit) but I am not necessarily aware of how my legs are moving, how I am balancing, the route I am taking, etc. Yet, perceptual information can trigger awareness of habitual aspects — the pain in my legs can make me immediately aware of how they are moving as I ride up the hill, seeing construction on the road ahead can make me immediately aware of my route. Once we become aware of the aspects, we sometimes cease to perform the whole behavior (or just that aspect) out of habit. I may intentionally slow down or turn right.

The second point is that both unintentional behaviors and actions can become habit. A child's first habits are habitual unintentional behaviors. A child who learns to focus his or her eyes, to move his or her arm in reaching, to crawl, etc. need not perform these behaviors intentionally, i.e. after an interval of suppression. One must be able to engage in an overt behavior by performing overt movements before one can suppress the overt movements. One must have a neuro-representation of that overt behavior in order to suppress it, since suppression arises from an association between a behavior and some other percepts. Thus, there have to be connections between neuro-representations of the behavior and other neuro-representations. Suppression itself is an unintentional behavior which becomes habitual. Most adults habitually suppress their object-oriented behaviors — we don't just go for things that look attractive, nor withdraw from things that seem unpleasant.
Actions also become habit. Consider a child learning to tie his shoe. Often he is being taught exactly what to do: first cross one lace over the other, bring the bottom one through and pull, then make one loop, bring the other lace around and up through the hole, and pull on the two loops. The child performs each step as an action: first he takes one lace in each hand, then he crosses them, then he tries to bring the bottom one through. The third step may be quite difficult, and unsuccessful, and much intentional behavior takes place as he learns to let go of the lace so that he can pull it through. The process continues. Eventually the child learns to tie the shoe. At first, the child will continue to perform each step intentionally. But, as the patterns of activity stabilize, the steps become habit. Rather than performing a series of small actions, he performs a single act of tying his shoe. The tying of his shoe is intentional, but everything involved in tying his shoe is habit. This act of shoe tying can also become habit. We habitually tie our shoes when we put them on, or habitually put on our shoes and tie them when we are getting ready to go out.

The third point is that the forming of the habit can also be an action or an unintentional behavior. Most habits are unintentionally formed. We just happen to do things over and over the same way or in the same circumstances and it becomes habit. Both habitual unintentional behaviors and habitual actions can be unintentionally formed. Suppression, holding one's head up, crawling, and using words are (for the most part) both habitual unintentional behaviors and unintentionally formed habits. Getting dressed, riding a bike, raising one's hand at school and saying please and thank you are (for the most part) habitual actions but unintentionally formed habits. Other habits are intentionally formed. For the most part we intentionally form habits that are habitual actions. My playing a piece of music is intentional (is an action), and so too is my practicing it over and
over. Often, intentional habit formation is also deliberate (a distinction that I will clarify later). For instance, I deliberately formed the habit of backing up my computer files, and saving early and often. Occasionally, one even intentionally forms habits of unintentional behaviors — I acted intentionally to form the habit of waking up at 6:00 (an unintentional behavior) by setting my alarm at six. Getting into the habit of using one's leg again after an injury, as one might do with the help of a physiotherapist, may also be an intentionally formed habitual unintentional behavior. Closely related to intentional and unintentional habit formation is the inducing of habits in others, primarily children. We can induce habits in others when we systematically and intentionally treat them or respond to them in certain ways. For instance, when we teach children how to perform an activity (setting the table, long division, tying their shoes, etc.) and monitor their practice, we induce habits in them.

Speech Acts

Many of a young child's early utterances are object-oriented in a way that might make them look like rudimentary actions. Consider the 'language game' that Wittgenstein describes in §2 (Philosophical Investigations):

Let us imagine a language for which the description given by Augustine is right. The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones; and that in order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out; — B brings the stone which he has learnt to bring at such-and-such a call.
— Conceive this as a complete primitive language. (Wittgenstein, Philosopheral Investigations: §2)

Wittgenstein's language game is modeled on early language use. On Augustine's account, a child's first utterances express desires and result in the fulfillment of those desires. As Augustine wrote, "...after I had trained my mouth to form these signs, I used
them to express my own desires." (Augustine, Confessions: 1.8) Indeed, many of a
child’s first utterances, such as “Mama!”, “Cookie!” and “Up, up!” seem to be elliptical for
the longer commands: “Come here Mama!”, “Give me a cookie!” and “Pick me up!”
But, as Wittgenstein indicates in his discussion of the builders’ language, there is
something wrong with conceptualizing these one-word commands as truncated versions of
the longer sentences since these longer sentences are not part of the speaker’s repertoire.
Similarly, there would appear to be something wrong with considering a child’s first words,
“Mama!” or “Cookie!” to be attempts to say, or easier-to-utter versions of, “Come here
Mama” or “Give me a cookie,” if these sentences are not within the child’s grasp. It is a
mistake to assume that children are just trying to do what more sophisticated language
users do with their words.

In the spirit of Wittgenstein, Austin and Searle, I am interested in determining what
children do with their first words, what behaviors or practices their utterances are a part of.
Like reaching or chasing, these one-word commands are unquestionably object-oriented.
The child wants something — an object, a person’s attention or presence, etc. But if a
rudimentary behavior is one in which the overt movements are suppressed prior to acting,
then these speech acts don’t appear to be candidates. After all, although children are later
expected to suppress some of their speech behavior (“Don’t interrupt! ‘Button your lips!’
‘Children should be seen not heard,’ etc.) there is no reason to think that any such
suppression occurs in these early speech behaviors. To begin with, even some of the
most impatient parents encourage early verbalization. In addition, there is no evidence (for
instance, partial suppression of lips moving, or whispering) that such young children are
suppressing speech behavior. But more importantly, the child needs to learn how to make
the words, perform the speech behavior, before he can suppress it. A child cannot
suppress what he/she cannot do. Only after a child starts to perform a behavior overtly, is it possible to suppress it. So, if these speech behaviors are to be considered rudimentary speech acts, it follows (on my theory) that something other than the verbal behavior must be suppressed.

Consider the following scenarios:

1. The child sees a biscuit and begins to reach for it.
2. The child sees a biscuit which would normally cause reaching behavior, but she is in the habit of suppressing such movements. She hesitates. Then, she reaches for it.
3. The child sees a biscuit and begins to reach for it. She can't get it. While attempting to get it, she begins to engage in an associated behavior, making the sound of the word, 'biscuit'.
4. The child sees a biscuit, but is in the habit of suppressing reaching behavior. She hesitates. Then she begins to reach. At the same time she begins to engage in an associated behavior, making the sound of the word, biscuit.
5. The child sees a biscuit, but is in the habit of suppressing reaching behavior. She hesitates. Then she begins to engage in an associated behavior, making the sound of the word, biscuit.

The first case is automatic behavior — the child immediately reaches for an object that she is attracted to. The second case is the rudimentary action — the child engages in the reaching only after a period of suppressed reaching. The third involves utterances, but no suppression or rudimentary action. But, in the fourth and fifth cases, like the second case, the initial reaching behavior is suppressed. Thus, it is possible for an action to be performed. However, the child does not just perform the overt reaching movement (which was suppressed). In four, the child performs the associated behavior of making the word \textit{in addition to} reaching, and in five, the child performs this behavior \textit{instead of} reaching.

In cases four and five, the speech behavior is oriented towards the same object to which the suppressed behavior is oriented. We saw earlier that the period of suppression creates an opportunity for new perceptual input, as well as an opportunity for selection (to
act or not to act). It now becomes apparent that the period of suppression also creates an opportunity for new, associated behaviors to begin, and thus an opportunity for selection between different behaviors. In other words, if the child is suppressing one object-oriented behavior and during the interval of suppression begins to engage in a second behavior oriented towards the same object, then there are two behaviors to select between.

Exactly how this works depends on the scope of the suppression. Consider the child who is performing three behaviors, reaching, uttering and suppressing. The suppression either serves only to prevent the movement of her arm, or it serves to suppress all her movements — she neither moves nor moves her mouth to make the words. If the former, then although the child would be suppressing the reaching, she would engage in the complete speech behavior, and say the word, ‘biscuit’, as in the fifth scenario. If the latter, the situation is slightly more complex. At least one of the three behaviors eventually has to stop. If the child stops the suppressing behavior, then she might perform both other behaviors, reaching and saying, ‘biscuit’, as occurred in the fourth scenario. In such a case, the reaching behavior is a rudimentary action, but it is accompanied by speech behavior. If the child stops the reaching behavior at the same time as she stops the suppressing behavior, since these two behaviors are closely associated, she will be left engaging in the unsuppressed uttering behavior — she will say, ‘biscuit’, as in scenario five. But, if the child stops the reaching behavior, she will be left engaging in the suppression behavior and the speaking behavior — the behavior being suppressed can change. Then, if she stops suppressing, she will complete the uttering, and perform a rudimentary speech act, again as in scenario five.

When new behaviors start to occur during the interval of suppression things become complex. I do not pretend to offer an argument that a child’s first words must be a
result of one particular process or another. My argument is that an interval of suppression can create opportunities for selection between behaviors oriented towards the same object. There are a lot of different ways in which this might occur, depending on the scope of the suppression as well as contingencies of the situation. But whenever two or more different behaviors are suppressed at once, then there is an opportunity for selection. Even if only one behavior is suppressed, the fact that it is suppressed can provide an opportunity for a second behavior to replace it.

Details aside, it seems reasonable to say that any overt, object-oriented behavior that follows a period of suppression is an action, whether or not the overt behavior was, in fact, the behavior that was originally suppressed, provided that the suppressed behavior is oriented towards the same object as the overt behavior. Thus, when young Augustine first begins to express his desires with words instead of reaching, he performs rudimentary speech acts. Suppression allows him to do one thing instead of the other; since both overt behaviors are real possibilities. Indeed, he performs both behaviors even though only one of them is performed overtly.

Suppression allows an opportunity for selection between like-oriented behaviors. When we learn language, one of the things we learn is to resist behaving in a directly object-oriented way (just going for it, reaching for it, chasing it) and to behave, instead, in a conventionally object-oriented way. Speaking, like other conventional object-oriented behaviors such as pressing buttons and turning knobs or handles, is only object-oriented in social contexts where an association between the behavior and the object in question has been created through convention and design. The conventions of the laboratory experiment account for why the rat's pressing a lever is food-oriented, just as the
conventions of the language-using community account for why the utterance, 'Biscuit', is object-oriented.

Conventions account for why any particular behavior is an object-oriented behavior, and thus, why one can come to believe that acting in that way will obtain the object to which the action is oriented. Conventions also account for why an individual might think one way of getting something (asking for it) is a better way than another (taking it, or demanding it at gunpoint), given a context of a community governed by convention. But, in order to perform conventional actions (including speech acts), one ought to be able to perform the conventional act instead of another behavior. Otherwise, it would appear that one cannot help but perform that action, rather than some other. This is out of keeping with traditional attitudes towards action. We don't just learn to say "Cookie" to get a cookie, we learn to do it instead of what we might otherwise do — grab the cookie and run.

The interval of suppression provides an opportunity for selection between competing object-oriented behaviors. This increases the range of possible behaviors in response to a single stimulus. If one is performing two behaviors simultaneously, suppressing both, then it is inevitable that some sort of selection will be made: one must perform both, neither, one or the other. I have characterized early speech acts as means-to-an-end selection process. A person begins to perform two competing but suppressed behaviors oriented towards an object, which makes it possible that one of those behaviors will be selected, and the behavior completed. In this way we learn to utter words, instead of doing other things, in order to achieve a desired end. I am not claiming that all such speech acts involve active suppression. Rather, my hypothesis is that in order for a child to begin to perform object-oriented speech acts, he must be capable of suppressing behaviors that are oriented towards the same object as the speech act.
As children gain skill in speaking, they acquire more sophisticated speech habits. They put words to different uses, not only commanding, but also greeting, scolding, disagreeing, thanking, accepting, labeling, categorizing, praising, asking, warning, inviting, apologizing, etc. Gradually a child forms associations between an utterance (and related actions) and the contexts in which it is appropriate. The child begins to say "Good-bye" and wave when people leave, to scold when something bad happens, to say no when they do not want something, to reiterate warnings near dangerous things. These utterances are not object-oriented in the same way that commands are, but they certainly have some focus or orientation — they are situated behaviors.

Children are taught, explicitly or by example, to behave in certain ways in certain situations. That they are so taught implies that they would not otherwise behave in just those ways in such situations. Thus, some situated speech behaviors are performed instead of other behaviors, in the same way that some commands are performed instead of reaching. A child begins to follow her father when he leaves the house. Instead of following, or perhaps wailing, the child is taught to say "Good-bye!" and wave. A child is playing and she drops something, instead of crying, or ignoring it (continuing her playing), she is taught to say, "Uh-Oh!" to focus her own attention (or someone else's) on the object. These speech behaviors, like commands, may be produced instead of other similarly situated behaviors if those other behaviors are suppressed. Suppression can allow for selection between competing behaviors in the case of situated speech acts, just as it does for object-oriented behaviors. And again, it is not necessary that every instance of a situated speech act involves suppression. Rather, the ability to suppress situated

---

5 Quite a few of a child's early uses of words are what Austin calls behabitives, which involve reaction to other people's behavior. This should be expected if people's utterances to children are also behabitives. And indeed, much linguistic interaction with children does involve reaction to their behavior. That is, of course, how we teach them how to behave.
behavior allows the child to develop the capacity to perform speech acts rather than other behaviors which might also suit the occasion.

This account of speech acts as situated verbal behaviors performed instead of other similarly situated behaviors is compatible with a Gricean analysis of speech acts (Grice 1957, 1969). On Grice’s account, also supported by Strawson (1970) and Bennett (1978), for a speaker to mean something with her words, she must intend to induce some sort of response in a listener — to get a cookie, to be picked up, make someone else aware of one’s feelings or beliefs, to make someone else believe something, etc. On my account, an overt behavior is intentional (and thus, an action) if the behavior first suppresses, or has the capacity to suppress, the overt movements that constitute that behavior. It is intentional if the behavior is performed instead of a similarly situated behavior. The capacity to suppress the behavior is closely related to desire for an object, a response from another person or for some other situated event. Thus, if a person performs a speech act instead of performing a similarly situated non-verbal action, then she intends to produce with her speech act the same desired effect as the similarly situated non-verbal action. For instance, when the child says, “Cookie!” instead of grabbing the cookie, she intends to produce the same desired effect as grabbing the cookie would produce. That is, she says “Cookie!” with the intent of getting the cookie. But unlike grabbing the cookie, saying “Cookie!” produces the desired effect indirectly by producing a response in the listener — the listener gives her the cookie.

As we become more sophisticated language users, we begin to use words to produce quite varied responses. For instance, we utter commands to induce others to

---

6 In addition, the speaker must have a few related intentions. The speaker intends that the listener recognize that the speaker intended to produce the response, at least in part from the utterance; and the speaker intends that the listener’s reason for responding in the intended way (if he/she does so) is at least in part the listener’s recognition that the speaker intended to produce that response. (Grice 1969)
follow the commands, we promise overtly so that others will expect us to perform the promised action, we tell lies to induce others to believe things, we tell tell jokes to amuse each other, we communicate our feelings, beliefs, hopes, fears, desires, and intentions so that others will be aware of our feelings, beliefs, etc. I am not claiming that every speech act involves the suppression of similarly situated non-verbal behaviors. Rather, we learn to perform speech acts by performing them instead of similarly situated non-verbal behaviors, other behaviors which would be performed with the same intention. We learn to produce desired effects or responses in other people by using words instead of acting without words — grabbing, reaching, crying, kicking, etc. Likewise, a child learns to perform intentionally some of these behaviors when she learns to do them instead of not behaving (following a period of active suppression), for instance when she reaches instead of stopping, and after a period of suppression.

Deliberate Action

The opportunity that suppression yields for selection between behaviors does not involve conscious selection or decision. Yet many of our actions and speech acts are deliberate, explicitly chosen, planned or decided upon. In such cases, we are well aware of making a selection. It is not reasonable to suppose that this type of selection between activities involves the suppression of the activities chosen between. When I try to decide which route to take to work, I don't usually begin going in both directions at once (in my head). When I plan my day, I don't begin to perform (but suppress) the activities I will perform later. When I resolve to call you, I don't begin to call you in my mind, suppressing the behavior until I am near a phone. Deliberate acts, of course, do require suppression in a trivial sense — if one didn't suppress immediate desires and always acted on impulse, one couldn't act deliberately. But, the more interesting questions concern the mental
behavior (consciously planning to, deciding to, intending to or meaning to do something) whose presence seems to distinguish deliberate acts from other actions.

Deliberate acts are, generally speaking, acts that one consciously resolves\(^7\) to do prior to doing them. First one performs a mental behavior of resolving to do something (Davidson\(^8\) calls this 'pure intending'). Then later, if all goes well, one performs a deliberate act that is importantly connected to that mental behavior — it is the act that one resolved to do. For example, suppose I deliberately come to visit you at ten o'clock on Monday. At some point on Sunday, I make that resolution. Then, I visit you at 10:00 on Monday, (or more precisely, I begin to prepare to visit you at 9:30, leave my house at 9:45, and arrive at yours at 10:00). If I do visit you deliberately on Monday, then whatever I do on Sunday has a causal effect on what I do on Monday. However, if I don't visit you on Monday, then my behavior on Sunday is to no avail, even though I truly intended then to visit you on Monday. Yet, the act that I perform on Sunday is the same act whether or not I fulfill that intention. This puts some stress on the causal connection between the two acts. What determines whether or not I act according to my prior intentions?

Some people are better at acting according to their prior intentions than others. Most people are acquainted with somebody who always follows through on his intentions, and with somebody else who seems to do it only rarely. The first is completely reliable (perhaps to a fault). The second means well but is, perhaps, absent-minded. He is like an eager five year old — whatever he meant to do seems to be cleanly wiped from his mind when the time comes round to doing it. Young children are rarely good at doing what they intend to do. Perhaps they are so poor at it that we cannot even justify the claim that they

\(^7\) Or consciously plans, intends, means or decides to do.
\(^8\) See Davidson's "Intending" (1978), reprinted in Davidson (1980). Anscombe (1959) refers to it as 'expression of intention for the future' or 'intention in a proposed action'.
intend to do it at all. At any rate, by adulthood most people can act according to their intentions, at least much of the time. Some are fully reliable, others less so. But doing what one intends to do is certainly something that one can get better at. This suggests that it is a skill or a habit, that it is something that practice improves. But what kind of skill is it? And how do we develop it?

Children often learn how to do things by mimicking others. However, forming an intention is primarily a mental act, and thus cannot be observed by others, making it difficult to mimic. Yet, it can be accompanied by an overt act such as stating one's intention. The fact that it is often accompanied by an overt act does not imply that statements of intent are merely descriptive of the mental act. Some statements of intent are the behaviors with which one forms the intention. When I say, "Leave the laundry, I'll take care of it later," I may be forming my intention by saying it. In contrast, when I say, "I intend to stop at the store on the way home", I am usually reporting on an intention that I have already formed. On this matter, I am in complete agreement with Austin. On the subject of commissives, he writes:

There is also a slide towards 'descriptives'. At the one extreme I may just state that I have an intention, but I may also declare or express or announce my intention or determination. 'I declare my intention' undoubtedly does commit me; and to say 'I intend' is generally to declare or announce. (Austin 1962: 157)

Thus, although children cannot observe, and thus mimic, people forming intentions (mental behaviors), they can observe people stating their intentions or describing their intentions.9

Children also observe some sort of correlation between commissive speech acts and later behaviors. "You can have a cookie later" correlates with getting cookies later.

---

9 To be exact, some statements of intent are also formations of intent (I form my intention to do the laundry later when I tell you that I will). There is often no way of telling the difference between a description of an intention that one has and a statement with which one thereby forms an intention. From the point of view of the observer, it rarely matters whether an utterance is a description or a formation of intention.
"I'll get you a glass of juice after we put away the blocks" correlates with getting a glass of juice after the blocks are put away. Thus, they learn to associate commissive speech acts with later, related behaviors. Indeed, it is widely accepted that it is vitally important that children learn this association early by observing their parents. One of the most common pieces of child-rearing advice is that you should do what you told your child you would do, and its corollary, that you should not tell your child you will do something that you are not going to do. 10 Thus, a child learns to associate a commissive statement with the later performance of the action committed to.

The child also learns that when a person commits him/herself to later action, he or she need not act immediately. Thus, a child learns to situate such speech behaviors. In response to, "Put away your toys, please," the child might say, "I will put my toys away later." She does this instead of doing what she would otherwise do (put the toys away immediately, wait, say "no!", etc.) Alternatively, if the child responds, "No!" or wails, it may cause the caregiver to say "Okay, but you must put your toys away soon." Thus, either the child or the caregiver can commit the child to later action with a situated commissive statement. Of course, there is no reason, at this point, to think that the child is conscious of committing herself to the later action; but she is nevertheless committed.

The last stage of learning to do what one consciously intends to do involves being held to one's word (or in the case of the caregiver's commissive, someone else's word). Sooner or later, the caregiver tells the child, "Now, it is time to put your toys away."

Although this stage of the learning process is not always smooth (since it can be difficult to

10 In fact, unlike may other pieces of child-rearing advice that go in and out of vogue, this one is omnipresent. There are disagreements about some of the details: whether it is okay to tell a child that you will do something that you aren't planning on doing if the child won't find out; what should one do when one fails to follow through; is it okay if one didn't explicitly promise; whether it is okay to make threats (of punishment) that one is not prepared to follow through on if the child obeys, etc. Yet all the discussion is against a background assumption that, in general, parents should do what they tell their children they will do.
induce the child to perform the act), the child does get held to her commitments. She
learns to do later (albeit, when commanded, reminded or forced to do) what she is
committed to doing by an earlier speech act. With practice, she becomes better at doing it
even without being reminded at the later time. (Yet, even a teen-ager needs to be
commanded occasionally, and most adults sometimes need to be reminded.) The child
develops the skill of following through on her explicit intentions. When the child learns to
think in words, that is, learns to suppress the overt muscular movements associated with
speaking, she can form intentions in her head just as easily as she can do it aloud.

This characterization of forming prior intention is verbal. Yet, even though many of
our conscious intentions are formed when we think to ourselves in words, "I'm going to do
that", prior intentions need not involve overt speech or verbal thought. Sometimes one
forms an intention by visualizing what one is going to do, imagining it, or rehearsing it, etc.

My claim is not that every formation of prior intention is verbal, but that we learn to form
prior intention through observing people's commissive speech acts and their later actions.

In general, learning to do later what one commits oneself to doing earlier requires that one
can observe other people making (at the time of commitment) some sort of overt sign that
represents the later action (that will also be observed) and that represents it as occurring
later; and likewise, it requires, at the time of follow-through, some sort of overt sign that
represents the same action to be done, but that represents it as occurring now.

Commissive statements and commands do this. Of course, there are many other ways of
doing it, for instance with pictures or actions (perhaps charades). But we generally do it
with words. Thus, our children first learn to do it with words.12

---
11 This account is in some ways similar to Anscombe's discussion (1959) of doing (later) what one
intended to do (earlier), which she compares to following an order or using a shopping list. Both the
order and the list can be issued either by another or by oneself.
12 It is easy to imagine teaching children to do it with pictures. A kindergarten class in our
community school had a wall on which there were photographs of the class areas (cut and paste,
A deliberate act, then, is an act that one performs after having consciously formed a prior intention to perform. We can form the conscious intent long before performing the deliberate act or immediately before performing the deliberate act. Both the act of resolving to do it later and the act of recognizing that it is time to do it now can be done overtly or internally. Following an act of resolving with an act of recognition (that it is time to do it) at the appropriate time and place is a skill, or habit, that one acquires through practice.

Acting deliberately is a skill or a habit that one acquires through practice. We learn to remember later what we intended to do at an earlier time. But, with any practiced skill there are many different ways that things can go awry — some people will just fail to acquire the skill, and even if one does acquire the skill, other things can interfere with the ability to perform the skilled act in certain circumstances. This conception of deliberate action as following from habitual, internalized and situated behaviors of resolving and remembering also provides a context in which one can understand the different types of failures to act as one intended.

Absent-mindedness, for instance, can be understood as a failure to develop the habit of remembering later what one intends to do now. This failure of habit may be connected to certain kinds of situations or circumstances. Many people habitually eat in different ways (according to different rules of etiquette) when with family, with guests, at

dress-up, water/sand play, legos, letters, numbers, etc.) Each photograph had a number of hooks below it — the same number of hooks as children were allowed to play there at once. For every child there was a card with his/her own photograph on it. During free time, they had to put their photo-card on a free hook in an activity area they were playing in. It is easy to extend this model by adding a time component. (Different sets of pictures for different times of day, or just ‘now’ and ‘later’ hooks.) At first, one puts the child’s photo on the hooks for him. One overtly holds him to his commitments (by showing him the picture at the appropriate time and having him perform that action). Later, the child will begin to put his own picture on different hooks, and one can hold him to these commitments as well. If he internalizes this behavior, his primary mode of forming prior intention, and acting deliberately, might be visual rather than verbal.
restaurants and alone. Thus, I might have acquired the habit of keeping my elbows off the table in all circumstances but the last. Likewise, a person might have acquired the habit of remembering later what he intends to do now in some circumstances (at work or at home) but not in others (for instance, social obligations).

In contrast to the absent-minded, a person might have become so skilled at successfully doing later what one intends to do now, that she begins to form more prior commitments than she can keep straight. She then does not successfully remember to do it all, at least not at the right times (since she can only do one thing at a time and then by the time she is finished one thing, the time for remembering another is past). Or, she might become scattered, trying to do everything that she intended to do by starting one thing, then interrupting it with another deliberate act, etc. Yet in general such a person has successfully acquired the habit of remembering to do things later. Intending to do things later can be quite a successful strategy if one develops the habit of remembering later, but it fails when over-used. This kind of failure in deliberate action also suggests the appropriate solutions: daily planners, agendas, calendars, etc.

Some forms of procrastination can also be understood as over-using the same strategy of forming conscious intent to do something later instead of doing it now. If I form an intention to do something later (and this a generally successful strategy for me), then when the time comes to do what I intended to do earlier, I might form a new conscious intention to do it later (instead of doing it now when I had earlier planned to do it). After all, forming prior intention is often situated such that one can form the prior intention instead of doing it immediately. This is a common way to procrastinate. And, again, the kind of failure also suggests the solutions. People who procrastinate in this way often look for ways to force themselves to perform the action — by asking someone else to impose a
deadline (imposing them for yourself is often part of the problem since it is a way of forming conscious prior intent), by permitting themselves a treat only if the task is complete, or by working in a place where there is nothing else to do (if necessary, even deleting games from one's computer.)

You might have acquired the habit of remembering later what you intended to do, but in addition you might acquire the habit of reconsidering earlier decisions, or changing your mind — thereby cancelling the prior intention. The person who acquires the habit of saying "My mind is made up" and refuses to reconsider, even though there is good reason to, may be protecting himself from developing the habit of changing his mind. If one reconsiders too much, then one can get out of the habit of just following through on earlier intentions. Likewise, making public declarations of intent, or promising, rather than just telling yourself that you will do it later, can help to limit the development of the habit of reconsidering.

Failure to follow through on intentions may be the result of a wide range of causes. Some of these are, of course, not philosophically interesting (such as physical inability due to injury or sickness, external physical prevention, etc.) But many of the causes for such failures of follow-through raise issues of philosophical and psychological importance. When an agent has already developed habits of following through, certain kinds of failure stand out as deeply puzzling. We attribute such failures to factors such as neuroses and other psychiatric dysfunctions (depression or more specific self-destructive tendencies), non-trivial absent-mindedness, different forms of procrastination, and of course the classic philosophical catch-all, 'akrasia' in its various forms. While I cannot explore any of these issues in this essay, the present chapter provides a general framework adequate for the analysis and description of such phenomena.
Chapter 7

DOING THINGS IN OUR HEADS

Actions follow from mental behaviors, at least for conscious beings. Some mental behaviors have direct consequences for the performance of overt actions; the things that we do when we act are causally connected to mental behaviors such as feeling desire, anger or disgust. In the case of deliberate action, we not only engage in a mental behavior but perform a situated mental act: we resolve to do something later instead of doing it now. If we successfully do such an action later, it is in part because we recognize then that it is time to do it, and in part because of the prior mental act of resolving to do it at such a time. The later performance is in this sense a consequence of at least two mental acts, the act of resolving to do it later and the act of recognizing the time to do it.

But these are not the only kinds of consequences that mental actions can have for overt behavior. Consider the relation between the following: a mental act of calculation and the overt act of giving the answer; a mental act of deliberation and the overt or mental act of doing something or resolving to; acts of mental rehearsal and overt acts of performance; a mental act of visualizing something and the overt act of drawing or describing what one has imagined.

In this chapter, I examine some of the different consequences of doing things in our heads for other actions, either mental or overt. Mental acts are, on this account, internalized overt behaviors, at least in origin. First we learn to do something overtly, then we begin to do those things 'in our heads' by suppressing the overt movements that would otherwise accompany the overt actions. So in order to understand mental acts, we need to consider how overt actions (our own or other people's) can have consequences for other overt actions.
When we make a promise or declare an intention, we are expected to do something later. Sometimes doing it later is no problem at all, especially if we are in the habit of doing later what we now intend to do. Other times, we must find some way to ensure (or make likelier) that we will do it later; otherwise we risk not following through on our intentions. People generally resort to such means when they think that they might otherwise forget; perhaps the intended act is too far in the future or there is so much to do that it is hard to remember everything. Perhaps one recognizes that one is absent-minded and that the risk of otherwise forgetting is quite high, despite one's good intentions. Or maybe the task to be remembered is too important to risk forgetting, no matter how unlikely it is that it will be forgotten.

The easiest way to ensure that we will do something is to engage in some sort of behavior at the time of resolving to do something that will result in the intended act. I could, for instance, just start doing whatever I intend to do at the time that I consciously intend to do it. That would be a convenient way of fulfilling my intention. But often one does not wish to begin the action immediately. Thus, there must be other things that we can do at the time that we commit ourselves to doing it later that will make it likely that we will indeed perform the action later.

Here are some of the things that I can do:

1) make a list of things to do.
2) write a note to myself.
3) tie a string around my finger, put a rubber band on my wrist, etc.
4) leave an object involved in the task in a place where I know I can't help but see it. For instance, I can leave the paper I said I'd read in the middle of my desk, or leave the pile of unfolded laundry on my bed.
5) ask someone else to remind me.
Some of these techniques have even become incorporated into commonplace practices. Consider the use that we make of daily planners, in-boxes and out-boxes, memo notes and boards, post-its, white boards pre-printed with "Things to do Today" or "Today's Chores", blackboards at school devoted to a list of today's work, homework books, etc.

But we also perform some of these acts in our heads. The verbal behaviors are the most straightforwardly internalized — people make mental notes, lists or plans, they tell themselves that they should remember (a mental counterpart of asking someone else to remind you). Yet we also perform mental acts akin to leaving the task in an obvious place or tying a string around a finger, which both ensure that some sign of the intended act will be 'in your face' at a later time. We form mental associations between the task, or some sign of the task, and other duties or objects that we will undoubtedly come across in the course of time. For instance, I might form an association between an errand I need to run and a certain corner that I pass by (where I could turn in order to run that errand). If all goes well, the thought of the errand will be triggered at the appropriate time and place, i.e. when I pass that corner.

When people perform the overt actions, we can understand (fairly well) what they are doing since we can observe them writing notes, making lists, sticking post-its in various places. We also understand, for the most part, why these techniques work. After all, we also see the notes, lists and such, and interpret them as signs of things to be done. This is especially true of verbal expressions of future tasks since we can read them or hear them; but we also recognize many non-verbal but conventional signs (the string around the finger, in-boxes, etc.) We also recognize when an item or task has been left in an obvious place (in the middle of the table or on the counter). In addition, we occasionally come to distinguish people's idiosyncratic signs of future tasks — Danielle turns her engagement
ring around so that the stone is on the palm side of her finger if she has something important to remember. However, it is not so obvious what exactly we are doing and why it works when we make a *mental* note or list, or when we form associations or use other mnemonic devices.

The suppression account of mental behavior implies that doing something 'in your head' is essentially the same as doing it overtly, except that (in the former case) the overt movements are suppressed. Yet, it is obvious that we don't do these types of things exactly the same way. When I write a note with pen and paper, I have to first locate the pen and pad, take pen in hand, remove the lid, then move my hand such that the words are written on the paper. I do not suppress all of these movements when I make a mental note. This is not what I am doing in my head. I suggest that expressions such as 'make a mental note', and other expressions which describe doing such and such 'in your head', 'mentally', 'silently' or 'in your mind', are used to pick out families of mental actions. Each family consists of mental actions which, for the most part, have the same effect, or are similarly situated, as the overt action corresponding to the expression. Thus, 'making a mental note' describes a wide variety of mental actions that serve much the same purpose as making a note (in that it is a similarly situated behavior). This family might include the internalized writing of the note, but it would also include visualizing a note, saying the words to oneself, forming associations between the task to be done and the time one must do it, imagining oneself doing it, etc.

Of course, most general descriptions of actions pick out a set of actions that share an effect: there's more than one way to skin a cat, fifty ways to leave your lover, and countless ways to spend money. Consider a few of the different ways to calculate the answer to 548 minus 257:
a) Say aloud: Two hundred from five forty-eight is three forty-eight, minus fifty is two ninety-eight, minus seven is two ninety-one.

b) Say aloud: Five forty-eight minus two fifty-seven. Seven from eight is one, five tens from four won't go, five tens from fourteen is nine, that's ninety-one, and two hundreds and the borrowed one from five six two, that's two ninety-one.

c) Say aloud: Пятьсот сорок восемь минус двести пятьдесят семь будет триста сорок восемь минус пятьдесят будет двести девяносто восемь минус семь будет двести девяносто один.¹

d) write down: \[
\begin{array}{c}
5 \ 4 \ 8 \\
- 2 \ 5 \ 7 \\
\hline
2 \ 9 \ 1 \\
\end{array}
\]

or e) write down: \[
\begin{array}{c}
4 \ 1 \ 4 \\
- 2 \ 5 \ 7 \\
\hline
1 \ 5 \ 1 \\
\end{array}
\]

f) use an abacus.

g) use a calculator.

These are all ways of calculating the answer. The way we perform the task depends on what we have learnt to do.² Since I have learnt to speak English, and not Russian, I can calculate the answer by the general method of speaking in English as in a) or b) but not by the general method of speaking in Russian as in c). Since I have learnt to write, I can calculate the answer with pen on paper as in d) or e). We also learn methods for calculating with the aid of instruments such as abacuses and calculators. Other differences may be better characterized as difference in technique, or using a different algorithm or problem solving method. For instance, a) differs in technique from b) but not from c). And d) and e) differ in technique from one another. But all are different ways of calculating the answer (rather than different instances of calculating the answer).

But, you can also calculate the answer in your head. Doing it in your head may, at first, look like a different general method of calculating an answer — you can do it aloud in English or in Russian, you can do it with pen and paper, or with an abacus, or in your head.

¹ have it on the authority of Inna Kupreeva, to whom I am indebted for this translation, that this is (more or less) the method of a) in Russian.

² Learning, again, just means that it is based on past experience. Being explicitly taught, figuring it out or bashing it out for oneself, copying someone else, etc. are all ways that we learn to do things.
But, doing it in your head is not really a different method of calculation: the methods may be the same except for the fact that the overt movements are suppressed. After all, most of these overt methods can be internalized. I can think the English words of a) or b); Inna can think the Russian words of c); you might visualize d) or e); I might imagine going through the motions of writing down d) or e); and a person who regularly uses one might even be able to visualize using an abacus. The phrase, 'calculating the answer in your head,' can be used for all of these behaviors as well as many others. It picks out a family of mental behaviors, such that each member can be understood as an internalized overt behavior of calculating the answer.

The fact that we cannot observe people doing things in their heads may lead us to think that people do what they do in their heads the same way. Different people can all do the same thing in their heads, in that they can all perform the mental calculation of 548 minus 257. But, they do not necessarily do this in the same way. We might say someone makes a mental calculation when she performs a mental act of any of the types described above. Since mental actions can be performed in such different ways, how we describe a person's mental action, for instance, as calculating the answer or as making a mental note, often depends not on the specifics of our behavior, but on the how the action is situated. I describe Inna as calculating the answer in her head because she performs a suitably situated mental behavior (one that I can neither do nor understand since she does it in Russian). I certainly don't do that when I calculate the answer in my head.

Of course, we readily accept some version of this for overt action. Although philosophers disagree as to whether a particular act of moving my arm can be described in many different ways, "moving my arm", "waving to the driver", "hailing a taxi", etc. or

---

3Of course, I can also imagine using a calculator. But although I can imagine doing exactly what I do when I do it overtly, I do not thereby calculate the answer.
whether each describes a different action; nobody doubts that any one of these
descriptions might be used to refer to an infinite number of other particular actions.

"Moving my arm" can describe a wide variety of acts, some performed by me (the motions
I make in raising my arm, throwing a ball, or waving) and others performed on me (the
motions that occur when you shake my hand, or when I catch my sleeve in the door).

Whether or not I describe Tony’s act as “hailing a taxi” depends less on the specifics of
his action (does he move his hand this way or that, does he yell something, etc.) than on
how the behavior is situated (is there a taxi present, does he see it, etc.)

Each of the following mental actions might be performed in a wide variety of
different ways: ‘making a mental note’, ‘remembering where you put something’, ‘willing
yourself to do something’, ‘psyching yourself up’, ‘deliberating’ or ‘listening to your
conscience’. Each different way of performing each different mental action can be
regarded as an internalized overt behavior. But, what makes each particular mental act an
act of the type ‘making a mental note’ depends less on the specifics of the action (what
words you say to yourself, or what you visualize, etc.) than on how the behavior is situated
(did you just make an appointment, did someone just ask you to buy a bag of milk, etc.)

Thus, what makes a mental act a mental act of a certain type is its relation to other mental
and overt behaviors (both your own and other people’s) and its relation to later behaviors
of your own.

---

4 See Anscombe's criticism (1979) of Alvin Goldman's argument that different descriptions describe
different acts. For instance, if I move my hand, I move the queen, I checkmate you and I scare the
fly, Goldman maintains that these always describe four different acts, whereas Anscombe argues
that they can be four different descriptions of one and the same act.

5 It does, of course, depend to some extent on the specifics of the behavior, since not just any
mental behavior can be described as making a mental note. When I was figuring out how to fit the
desk in the basement into the back of a car, I was not making a mental note. There are lots of ways
to skin a cat, but what makes them all ways of doing one thing is a certain unity of function and
situatedness.
I 'make a mental note' in certain kinds of situations: those in which I need to do something later. When I make a mental note it usually has the consequence that I later do what I made the mental note to do because I realize later that it is time to do that which I made the mental note to do. My behavior in realizing that it is time to do it is often an habitual behavior. If I frequently write notes to myself or use a daily planner, then it is useful if I am also in the habit of reading those notes or consulting the planner later. Of course, some people who write notes or use planners do not get into the habit of looking at them later; instead I might get into the habit of performing an intentional or deliberate act of remembering what I wrote down earlier. In most cases, I do this instead of looking at the notes (perhaps due to disorganization or laziness or both) and gradually it becomes a habit. Similarly, if one makes a mental note, then one must perform some sort of behavior later that consists in remembering what one made a mental note to do. But, however you do it, writing notes only works if you do something later that will result in knowledge of what you wrote, and often, what you do later is an habitual behavior.

This habit might consist in exploiting temporal or situational triggers: first thing in the morning, when making plans for something else, at meals, before leaving a room or a building, when you change activity, when you get to work, before you leave work, when you get home from school, etc. These are all times at which you might be in the habit of consulting a daily planner, notes, or trying to remember the things that you consciously intended to do. Chances are good that you have often been asked to perform some action of this type at certain times or in certain situations. Your parents asked you, when you got home from school, what homework you have to do. Your spouse asks you, when you sit down for morning coffee, what are your plans for the day. We internalize these behaviors, and these kinds of situations become automatic triggers for the habitual
behaviors of consulting one's planner or recollecting one's plans. People also get into the habit of 'planting' triggers — you put the planner in your pocket or purse so that it is often noticed; you put a string around your finger; you put your list on the fridge door or on your mirror. This, too, can be internalized when we deliberately form associations between ideas: for instance, if there is something I need to do tomorrow I form an association between it and some routine morning activity (shower, coffee, e-mail check, etc.) then when I perform that morning activity, it will trigger the thought of what I have to do. In this way, performing a mental action such as making a mental note can have consequences for future overt behavior by creating a trigger for later behavior in much the same way that similarly situated overt actions do.6

Such acts as these (whether mental or overt) prepare us for future behavior. Indeed, many mental actions, as well as the similarly situated overt actions, prepare us for future overt actions in one way or another. Deliberation is often preparation for action or a resolution to act. Calculation, whether mental or overt, is often preparation for using the answer, for instance, in paying a bill, answering a question, or in deliberations about what to buy (what is the best deal). Visualizing is often preparation for such things as moving furniture, painting, deliberation about what to wear to a party, etc. We learn to do most of these things within the context of overt practices: we see other people calculating, deliberating, and drawing figures overtly; we see them perform overt actions that seem to

---

6 Mnemonic systems involving both mental and overt behaviors have been used since classical times for improving performance in later tasks that require either explicit or implicit recall. As Baddeley writes, "Mental imagery has formed one of the cornerstones of mnemonic systems through the ages." (1990:186) Yet, while mental imagery is an important way of forming associations, there are vast number of other types of mnemonic systems, including meaningful verbal associations, verbal systems that draw on first letters (such as "Every Good Boy Deserves Fudge" to remember the lines of the staff in the treble clef), systems that use rhythm and rhyme (such as the poem "Thirty days hath September..."), or methods such as elaboration by telling a short story to explain an association. For further discussion of the success of mnemonic devices see Baddeley 1990, especially chapter 8.
require such a mental act, and perhaps after the kind of pause that looks as if they are performing such an act.

In these types of situations, the preparatory action is a different type of action than is the resulting act. Calculating and announcing the answer are different types of action; as are deliberating and resolving; and visualizing and moving furniture. Yet, some preparatory actions prepare you for future actions by being the same type of action—practice or rehearsal. By playing a piece over and over, Ilona prepares herself to play the piece again, albeit in a different context (one of performance rather than rehearsal). By practicing a routine of song and dance, Tom prepares himself to perform it on stage. Practice can, in fact, be considered a type of habit formation (often, but not always, deliberate). One result of habit formation is that once a type of action becomes habitual, the performance of such an act requires less attention (at least to the aspects that have become habitual). Thus in performance one can focus attention on different aspects of the action than one does in rehearsal.

Practice and rehearsal can also be done in your head; you can prepare yourself in much the same way for future overt behavior. Before speaking aloud in a meeting, a person might ‘gather his thoughts’ by mentally rehearsing what he is about to say. Thus, when he speaks he can focus more of his attention on delivery rather than on his choice of words. In dance classes, when waiting for a turn to perform a dance routine, a dancer may go over the steps ‘in his head’. Perhaps he says the words that his teacher says when teaching the routine, “left, right, front kick, behind, turn, ...” Or perhaps he imagines (kinesthetically or visually, rather than verbally) performing these steps in sequence. Thus, when he performs the routine, he is better prepared to perform the steps in sequence, focusing more of his attention on other aspects of what he does (position of his body, etc.) Unlike overt practice, mental rehearsal does not train the muscles, nor does it
provide kinesthetic or other sensory feedback (preventing improvement of such things as being in tune when playing a piece, or one's balance or position in turning, etc.) Yet, the ways in which mental rehearsal does prepare you for future performance are otherwise quite similar to that of overt rehearsal.\(^7\)

Since mental rehearsal consists in internalizing overt actions and doing them in your head, the subsequent performance of the overt action which one has been practicing can be understood as re-externalizing the action. Of course, the context of the performance is very different from the context of rehearsal, and one generally focuses attention on different aspects of the action in the two contexts, and thus, in some ways the actions that constitute rehearsal are a different type of action from those which constitute performance. Yet, the difference between the two is not of the same order as that between calculating and answering a math question, or deliberating and resolving to do something. Unlike these types of preparatory actions, rehearsal prepares one for future action in so far as the future action is the same type of action as those that one rehearsed. We re-externalize the mental action when we do overtly the same kind of action that we did in our heads.

Except in the case of mental rehearsal, our mental actions do not result in the re-externalization of the type of action that we do in our heads, despite the fact that mental actions are, for the most part, internalized overt behaviors. The consequences of our mental behaviors are much more varied. Indeed, the consequences of any preparatory action (whether rehearsal, deliberation, calculation, etc.) go far beyond the performance of future actions which are thereby prepared for.

---

\(^7\) There is extensive literature on both overt and mental rehearsal and their effects on future tasks. There is some evidence that there are several different types of rehearsal, for instance maintenance and elaborative rehearsal, see Craik and Lockhart (1972) and Craik and Tulving (1975). Different types of rehearsal may have different effects, improving performance on different kinds of recall tasks. For further discussion see Baddeley (1990), especially chapter 7. For discussion of rehearsal through visual imagery, including mental motor rehearsal, see Kosslyn (1994), especially chapter 10.
I am particularly interested in what kinds of consequences might ensue from different methods or ways in which we perform an act. Hailing a cab by waving, by yelling and by throwing oneself in front of it may all be effective preparation for riding in a cab (they may all be effective ways of hailing a cab); but there may be other consequences in which they differ. Yelling may attract someone else's attention, my waving hand may hit someone or be mistaken for a greeting, and throwing yourself in front of the cab can be downright dangerous. The different ways of performing mental actions may also have consequences beyond the success or failure to adequately perform the acts which they thereby prepare you to perform.

Consider different ways in which I might remember where I left my keys. In many situations, for instance when I am leaving the house, remembering is a voluntary action. Ebbinghaus (1884, translation of 1885) distinguished between voluntary and involuntary remembering, opting to focus his attention on involuntary memory, as did most psychologists after him. Yet, some recollection is voluntary — for instance, when we are asked to recall facts or episodes from our personal past history, or when we actively try to recall where we left something important.8 Consider an incident of voluntary recall of the latter type: I need my keys, I don't immediately know where they are, and instead of looking for them, I intentionally try to remember where I left them. I can do this in many different ways. I might do this by recalling what I did after I last came home and used the keys to open the door. If I can remember the things I did, in sequence, then I might thereby remember doing something with my keys. Alternatively, I might start thinking about all the places where I usually leave my keys, asking myself, for each location, do I

---

8 We can distinguish both between voluntary and involuntary remembering in the sense of committing something to memory, and voluntary and involuntary recollection or remembering in the sense of retrieving something from memory. It is the latter type of remembering that I am interested in here. Such voluntary recollecting also depends on involuntary remembering in at least the first sense of learning or committing to memory, and may also depend on some involuntary recollecting.
remember putting them there? Or I might start thinking about where I would look for my keys, where I might have left them. I may run through a number of different possibilities, refining the search procedures until I am confident that I could find them quickly. Alternatively, I might try to form a visual image of where my keys are, particularly if I remember seeing them somewhere and it is just a matter of remembering where. I might try to remember where I told myself I was leaving my keys when I put them down.

What determines how I go about remembering where my keys are? It is unlikely that I decide how to remember where I left my keys. Chances are that I just start doing it the way I usually do it — according to my habit. Habits are generally acquired in social contexts. Consider the different sorts of things one might tell a child who cannot locate something. One might ask, "When do you last remember having it? What did you do next? What did you do after that?" If a child internalizes this behavior, then he may form the habit of trying to remember where he put something by recalling everything he did, in sequence. Or one might say instead, "Usually you put it on your bed, or on the window sill or in the front hall cupboard. Have you looked there?" If one internalizes this behavior then he might be more likely to remember by running through the usual places. If one tells the child to go and look for them, or if one helps the child look for them, then the child may internalize this strategy, mentally searching for the objects. Alternatively, one may tell the child that in the future he should notice where he puts things, to be more observant of his own behavior. If the child follows this advice, he may try to form clear visual images of where things are, or tell himself where he is leaving things, or what he is doing. Thus, he may later remember where he left things by recalling the visual image, or recalling the words he said to himself.

Of course, most children are taught a combination of techniques. They develop the habit of remembering one way or another, depending on how successful the strategy is for
them, or how often they are encouraged to use one strategy rather than another. Other features of their environment may also affect how successful the strategy is — whether the child is in an environment where everything has a place, whether there is a regular daily routine, whether there is always a lot going on, or whether things are generally left where they are, moved or put away by others, etc. In addition, a certain method of remembering may be successful for some situations, and other methods successful for other situations.

Remembering where one left something using any one of these techniques will often result in successful recall of where the object is. However, the technique used may also have other effects. If I am in the habit of reviewing what I did in order to remember where I left something, then my 'memories' of what I did are reinforced (rightly or wrongly, since I could have made mistakes when I ran through my mind everything that I did). But, in general, I may develop a better episodic memory. Thus, if I am asked, on another occasion, and for other purposes, "What did you do yesterday?" I might be quite good at giving an answer, since I am quite practiced in recalling in this way. On the other hand, I might be more likely to remember those events that I have already run through in my mind, whether or not they are the events that I am trying to remember.

If I develop the habit of noticing where I put things or what I am doing, then I may be, in general, more self-conscious of my overall behavior. I may also become more organized in general, since noticing where one puts something may lead one to notice that one is putting it in a dumb place and then to move it to a better place. In contrast, if I am in the habit of running through search strategies, I may be developing other skills such as problem solving, "Where would I have left them in such circumstances?" But if one becomes good at this type of problem solving, there is less of an advantage to being organized in general. After all, if I am good at figuring out where I left things, then it
doesn't matter if I am conscious of where I leave things, making it less likely that I notice that I am leaving something in a bad place.

Since remembering where you left something is a mental act, others cannot observe how you are doing it. Thus, it may seem as if everyone is doing it the same way, since, for the most part, we all do it in the same sorts of situations and get similar results. Yet, it is quite possible that different people do it in different ways, depending on the kinds of behaviors they have internalized. Indeed, many mental behaviors that have caught the attention of philosophers can be better understood by considering the different ways in which we perform mental actions, and the different overt actions which we might have internalized to make such mental actions possible.

Consider the different ways in which we might will ourselves to do something that we are afraid of doing. A man is terrified of heights, but must cross a high bridge to meet his lover or risk losing her. How does he force himself to cross? Through an act of will. It is useful to ask, what is he doing when he wills himself across the bridge? If he does will himself to cross, he must be doing something or another — and there are countless things that he can do. He might focus his mind on his loved one, visualizing her face, and telling himself, with every step, that he must reach her. He might focus his attention on the movements of his body, aware at every instant exactly what is under his feet or of the position of his body. Perhaps he coaches himself, 'Move this leg forward, now the other, hold on to the railing, reach forward, don't look down ...' He might imagine what his lover or other people would say if he failed, or if he succeeded. He might imagine the voice of his lover or his mother, encouraging him on, 'You can do it, I know you can.' He might say to himself, 'I think I can, I think I can.' He might try to imagine that he is just walking along the sidewalk.
All of these techniques are ways in which we might will ourselves to do what we would not otherwise do. The particular way in which we do it depends on what methods we have learnt, often from hearing people encouraging us to do something that we (or other people) are loath to do. In addition, there are countless books or stories that explain or demonstrate such techniques — self-help manuals, motivational tapes, familiar stories for children, such as *The Little Engine That Could*. We internalize the methods which work for us, even though none works in every instance.

If we get into the habit of willing ourselves to do something according to some particular method, or methods, then this may have implications for the kind of agents we are. Indeed, each method suggests a different conception of agency. Focusing attention on the goal implies that the mark of agency is striving for a desired end. Focusing attention on the movements of the body suggests that basic actions (in Danto's sense) are the locus of agency. Coaching oneself suggests that agency lies in reason, which is distinct from the body and can command it. Alternatively, agency may be linked to other people's expectations, self-confidence, or even self-deception. How we perform acts of will, or fail despite the efforts, may affect how we act as agents, and thus, how we see ourselves as agents. The converse may also be true: how we characterize ourselves as agents may constrain how we will ourselves to act, thereby reinforcing a conception of self-agency.

But in general, the kinds of agents we are and our conceptions of agency are dependent on the methods by which we perform mental actions — techniques that we internalize.

I do not mean to imply that whenever we perform such an act of will we begin to reflect about the kind of agent we are. We may not even be aware of what we are doing when we will ourselves to act, let alone how it effects our conception of agency. When we get into the habit of performing an action one way or another, we are often not conscious
of how we are doing something. Once we get into the habit of walking to the store by a

certain path, we stop thinking about the route we take when we take it. Thus, we aren't

necessarily conscious of how we perform mental actions. The point is that when we do

things in our heads, we do things in different ways, whether or not we are conscious of

them. How we perform them can have far reaching consequences for our conception of

agency, and indeed, for our conception of the mind.
I have presented the outline of a theory of behavior which is a theory of the development of behavior over a life. People are born moving — indeed, behaving, since their movements are self-initiated. They are born perceiving; their interaction with the world is sufficient for conscious experience. By considering the human being as a medium of representation we can understand a person's conscious interaction with the world as the cause of neural representations of the objects of their experience. These neural representations are utilized in the production of new behaviors and new ways of interacting with the world. As we experience correlations between the various objects of our perception, including perceptions of our own behavior, we begin to associate various perceptual events with one another. As behaviors increasingly utilize neuro-representations they become increasingly systematic, in so far as they develop in accordance with past experience of such correlations. Utilization of such associated neuro-representations also serves to reinforce the connections between them, and thus the most stable of associations are those which have been utilized most often. In this way we develop stable behavioral habits in accordance with complex neuro-representations which serve to organize our perceptual experience.

In a social context, the potential for correlation between events is greater, since different perceptions of objects correlate not just with each other (the sound and the look of a dog), but also with our perceptions of other people's behaviors, especially speech behaviors (the look of the dog and the sound of the word 'dog'). Thus the child's responses to objects that she perceives are in accordance with both her past direct perception of the objects and also her past perception of words or behaviors which
correlate with those objects. An object can trigger two contradictory responses — one which stems from her past, direct experience with objects and the second which stems from her interaction with other people with respect to such objects. For instance, a child begins to reach for something that she has learnt not to touch (from her parents' behavior towards her with respect to such objects). In such a situation, the neuro-representations of an object are utilized in the production of two conflicting behaviors, one which produces overt behavior and one which suppresses the overt movements involved in that behavior. This creates an interval in which the child is engaged in conscious reaching behavior but during which her overt movements are suppressed.

Suppression allows for the experience of distinction between perception and behavior. Without suppression, perception and behavior are one — the experience of moving is indistinguishable from the simultaneous perceptual experiences that one has when moving. When the overt movements are suppressed, one experiences something akin to moving without simultaneously experiencing the movement and the changing perceptions one normally experiences when moving. This experience without overt movement is mental behavior. In this case, it is the mental behavior of desiring, an object-oriented mental behavior that is closely associated with certain movements towards the relevant objects (reaching or grabbing).

When overt movements are actively suppressed, the child is in suspense — actively engaging in an object-oriented behavior and actively suppressing the movements that the behavior causes. Yet, she cannot continue in this state indefinitely; eventually one of those behaviors must cease. This creates an opportunity for selection. The interval of suppression creates a temporal gap between the start of the object-oriented behavior and its completion in overt movement. The gap provides an opportunity for receiving new
information, or engaging in new behavior, that can push her one way or the other — to go for it or to stop being oriented towards the object. She must either stop suppressing and allow the movements to follow, or cease the object-oriented behavior altogether.

When the child goes for it, reaches instead of stopping, she performs a rudimentary action. Her action follows a mental behavior of desiring. Indeed, those overt movements are caused by that very mental behavior of desiring. The mental behavior of desiring is, after all, the reaching behavior that normally results in overt movements, only with the overt movements suppressed. But when the child reaches after the interval of suppression, she acts instead of not acting. She has the capacity not to act, as is evident from her suppression. The suppression allowed for selection. Whenever a person does one thing rather than another, she acts. Although it is not the case that every action follows a period of active suppression, the capacity to suppress is necessary for action.

The capacity to suppress distinguishes the actor from the mere behaver. But in a conscious individual it also produces mental behaviors and mental actions. People do things in their heads. They internalize overt behaviors and actions — whether reaching or grabbing, recolling, talking, coaching, reading, drawing pictures, etc. They begin to do things in their heads instead of doing them overtly. In such situations we perform mental acts which would serve the same function as a similarly situated overt behavior.

Not only do we perform mental actions instead of overt actions, but many of our overt actions also involve the performance of mental actions. We sometimes act deliberately by first forming an intention (in our heads) to act in that way. Or we perform an overt act which we will ourselves to do — and willing can be understood as a mental action which accompanies our overt behavior. Or we pay our share of the bill after calculating in our heads how much we owe. But mental acts, like overt acts, can be performed in many
different ways. Different acts of ‘willing’ or ‘calculating’ may have little in common other than that they are similarly situated mental acts. But every time we perform a mental act, we must do it one way or another.

The way I perform a mental act will depend, for the most part, on which overt behaviors I have internalized. Thus mental actions reflect social conventions and social practices just as our overt behaviors do. When we think with words, we use the languages that we have learnt to speak. When we perform mental arithmetic, we generally use the kinds of methods that we have learnt to use overtly. When we commit things to memory and voluntarily recall them, we internalize the overt behaviors that we have learnt from others and which have worked for us. Once we start to do things in our heads, internalizing overt behaviors, we can begin to do new things, even performing mental acts that we have never (perhaps could never) perform overtly. Although these behaviors are not internalized overt behaviors, they are dependent on and constrained by the overt behaviors that we have internalized.

Our behaviors, whether mental or overt, are representational in so far as they utilize representational neuro-structures. Thus, the behaviors are interpretable. Given our understanding of the human being as a medium of representation, we can infer from our observations of people's behavior to the causes of their behaviors, that is, the kinds of past experiences which would account for their behaving as they do, given an assumption of shared behaviors. What Wittgenstein said about our interpretation of an unknown language is true also of our interpretation of novel behavior, “The common behavior of mankind is the system of reference by means of which we interpret an unknown language.” (1968:205) Thus, when we want to understand what people do in their heads, we must look to the kinds of things that people do overtly — overt behaviors which the mental behaviors are performed instead of; overt behaviors which have been
internalized; and overt behaviors which also incorporate or are caused by other mental behaviors.

We perform mental actions and develop mental habits in accordance with the ways in which we do them overtly. Yet many of our overt behaviors and actions are better understood as practices, intimately connected with institutions or organizational structures, such as churches, schools, political systems, economic systems, legal systems, sports, clubs, medical and psychological institutions, etc. Characterizing behaviors as practices focuses attention on how that behavior plays a role in a social structure. In order to interpret such behaviors, one must have an understanding of the social framework in which these behaviors develop. Like their overt counterparts, many mental behaviors or actions are best understood as internalized practices, in that they can only be understood within the context of the social framework in which they develop and occur. Some mental behaviors are learnt within institutional contexts and are internalizations of behaviors which are meaningful only with reference to that context: psyching oneself up or out, religious or spiritual meditation, silent prayer, silent reading, mental arithmetic, budgeting one's time, etc. Other mental behaviors or techniques for performing mental actions are acquired within a context of more general practice: thinking in English, or in Russian; promising oneself; making a mental map; deliberating by weighing the pros and cons, or by considering different means to an end; willing oneself to act by coaching, or by thinking about what other people might say; planning future actions by imagining what something will look like, or by giving oneself advice about what to do; fantasizing, etc.

An investigation of how we perform mental actions can provide some insight into the nature of the mind. The work I have done suggests that the capacity to act (to do one thing instead of another) and the capacity for mental behavior are intimately connected.
The capacity to suppress overt movement, and thus to do things in your head, is what makes possible voluntary overt behavior. All animals are born with the capacity to behave; to be a moral agent something additional is required. I have suggested that it is the further ability to resist behaving through the active suppression of overt movement that makes us agents. The capacity to suppress overt movements allow us to substitute a different (even novel) behavior that is better (or less) suited to the context than a similarly situated behavior which we might otherwise perform in such circumstances.

Furthermore, the relation between the brain and mental behavior is much like the relation between the brain and overt behavior. The brain is a medium of representation. Structures in the brain represent those things which have caused them in our conscious interaction with the world. Those representational neuro-structures are utilized in the production of behavior, mental or overt, which is also representational. The utilization of those neuro-structures involves much the same type of neural activity as did the forming of those neuro-structures, and are similarly sufficient for conscious experience.

Consciousness is a feature of vertebrate behavior precisely because it is a feature of vertebrate sensation of the world. Although not all behavior is conscious behavior, people are always conscious when they are learning, or forming new representational neuro-structures (although they need not be conscious of what they are learning). When people perform habitual behaviors they are often not conscious of all details of what they are doing (although, for the most part, they are conscious of something or things). This is also true of habitual mental behaviors. Awareness of what one is doing is often present when one learns a behavior which later becomes habitual. However, once a behavior becomes habitual, one need not be conscious of how one solves a mental math problem or of how one is deliberating, any more than one need be conscious of how one is riding a bicycle.
This considerably limits the utility of introspection as a means of understanding what we do in our heads.

The different kinds of relations between mental behaviors and overt behaviors are quite similar to the different kinds of relations that might hold between different overt behaviors. Overt behaviors prepare us for other overt behaviors in different ways: consider rehearsal, habit formation through practice, deliberation for future action, commissive speech acts, calculations for use in future deliberations or actions, drawing diagrams for future reference, making lists or notes, etc. Likewise, mental behaviors prepare us for other mental or overt behaviors: consider mental rehearsal, mnemonic techniques, mental deliberation for future action, commissive mental acts, mental calculations, mental imagery, making mental lists or notes, psyching oneself up, etc. Other behaviors, mental or overt, prepare us for future behavior in less direct ways: consider relaxation techniques, methods of amusing ourselves, ways of psyching ourselves up, prayer, resolutions, etc. Thus, questions about mental causation can be recast in terms of the causal relations between the brain structures and activities and behaviors and also in terms of the causal relations between different behaviors, whether mental or overt.

But what about the nature of the mind itself? Although I have presented a medium theory of representation, I have not characterized the mind as a medium of representation. Indeed, I have no notion of what kind of a medium the mind could be, how objects or events could causally interact with it, and what such alterations to its structures would consist of. However, behavior, including mental behavior, can function as a second-order medium of representation — second-order because the representational capacity of behavior is not due to interaction with that which it represents, but due to the fact that behavior is produced through the utilization of representational neuro-structures. Thus,
although the mind should not be considered a medium of representation, mental behaviors are representational, and thus interpretable, just as overt behaviors are.

Although there is little to conclude about the nature of the mind, this investigation has many implications for the nature of mental behaviors and other events. Rather than talk about the mind, we can think of all those mental behaviors and mental actions, conscious experiences, perceptions and other mental events as constituting a mental life. Just as your life is constituted by all that you do and all that happens to you, your mental life is constituted by all that you do and all that happens to you in your conscious experience. But people often do things in accordance with the ways in which they have done them before — out of habit. Thus, we begin to see patterns of behavior, or ways in which they tend to behave. We can speak then of the character of a person's life or the forms that it takes. These are the general behavior patterns of which particular actions are a part, and against which those actions are understood. Kevin climbs the glacier because he is an adventurer, and his being an adventurer consists in performing such activities as climbing glaciers, rock climbing, extreme skiing, etc. Another person might be described as a traveler, or a homebody, or perhaps a loner, social butterfly, family man, working girl, career woman, couch potato, confirmed bachelor, lady's man, gourmand, etc. The forms of a person's life are comprised by the activities that he or she performs. Of course, a person's life consists of countless patterns of behavior — and the dominant forms change through time and in response to changing circumstances. When we interpret what people do, we often look at those characteristic patterns, seeing the behavior as a part of the pattern (or as a full break from it, which can also be part of a characteristic pattern of behavior).
A mental life is likewise comprised of patterns of mental behavior. Thus, it takes on certain characteristics or forms. The things we have done in our heads, and the ways in which we have done them, influence what and how we will do things in the future (both overtly and mentally). We perform every mental act one way or another — what we do (what words we speak in our heads, what images we imagine, what motions we suppress) is part of a mental activity, or form of mental life. There are many different forms of mental life. Many people spend much of their mental life engaged in verbal thought of one kind or another. Perhaps I spend a lot of mental activity conversing with myself or others (in my head); or deliberating about means to an end in an Aristotelian fashion; or rehearsing what I will next say aloud; or constantly analyzing or rehearsing what I have said and done; or imagining what others would say. For others, less verbal forms of mental life might dominate — manipulating images in one’s head; projecting oneself into different situations; or fantasizing or visualizing; trying out different roles (in one’s head) in order to decide how to behave; or imagining how something would feel. The ways in which we perform mental behaviors, the particular techniques or methods that we internalize, have consequences for other things that we do — in our heads or overtly.

The ways in which we perform mental actions, the forms that our mental life takes on, have consequences for future mental and overt behavior and on our conception of the mind and ourselves as agents, persons, etc. If I frequently solve problems by imagining myself in conversation with others, then my mental life takes on a certain form. Thus, often when I try to will myself to do something that I am loath to do, I am quite likely do so by talking (in my head) to others — I imagine what advice they would give me about getting across the bridge, and I imagine talking to them about crossing after I have done so. I

---

1 "Here the term "language-game" is meant to bring into prominence the fact that the speaking of language is part of an activity, or form of life." (Wittgenstein 1968: 23)
might conceive of agency as social — to act is to change one’s position in an intricate net of overlapping social relations. I might conceive of the mind as a kind of inner social environment. On the other hand, if I spend a lot of time fantasizing, then my mental life might take on a very different form. In this case, I might will myself across the bridge by imagining that I am not walking across the bridge at all, but on the street or in the valley. I might conceive of agency as a form of self-deception — we trick ourselves into believing that we act, since the hurdles that we apparently face are merely a consequence of how we look at the world. I might conceive of the mind as a mechanism for generating different possibilities, different narratives, and different reconstructions and interpretations of the external world with which we interact. My mental life might take on either or both of these forms, and it might take on other forms instead or as well. A mental life consists of many forms, of countless different patterns of mental behavior. But the kinds of agent we are, the kinds of mind that we have, will depend on what things we do in our heads and how we do them.

We can interpret our mental behaviors in light of the patterns of mental activity or forms of mental life of which they are a part. But mental behaviors are also acquired through the internalization of overt behaviors. Thus, the dominant forms of a person’s mental life can also be understood in light of overt behaviors which can be internalized. Thus, within a shared social environment there is likely to be some convergence in the dominant forms of different people’s mental lives. However, this does not mean that the dominant patterns of a person’s mental behaviors are comprised of internalized activities that constitute the dominant patterns of his overt behaviors. A loner may spend much of his mental life in ‘conversation’ with others, and a non-traveler may spend much of his mental life fantasizing about being in other places. It does suggest, however, that the
dominant forms of a person's mental life are probably based on overt behaviors prevalent in that person's society.

Since we internalize overt practices, what people do in their heads varies from culture to culture, indeed, between families, institutions, generations, etc. The prevalence of different overt practices varies between social groups. Thus, what people do in their heads and how they do them will change through time and from place to place. This might seem to imply that my great-great-grandfather couldn't have done in his head some of the things that I do in mine. Of course, it is obviously true that Ezekiel couldn't plan how best to use his air mile points for a vacation trip, couldn't visualize the CN tower, and couldn't replay in his mind what he saw on television last night. But, there is a sense in which he could have done what I do when I perform mental activities such as psyching myself up or working out the bugs in a plan. Perhaps Ezekiel couldn't 'psych himself up', since this is a more recent practice, but no doubt he could tell himself the kinds of things that I tell myself when I psych myself up. He couldn't 'work out the bugs in his plans', since this is an activity modeled on computer programming, but he could run through variations of a plan in his mind in different ways in order to work out the problems. Yet, although he could do these things, there would be no reason to expect him to perform such mental activities. Unlike me, he did not live in a society in which people frequently psych each other up and work out bugs. Thus, people in my society are much more likely to internalize this behavior than people in his.

In this essay on mental action, I have presented a theory of the development of behavior, mental and overt, over a lifetime and within a society. This theory suggests that we can benefit from taking a slightly different approach to understanding mental phenomena than those which have been adopted by other contemporary philosophers of
mind and action. Questions such as, "What is the nature of the mind or consciousness?" or "What is the relation between the mind and the brain, or between the mind and behavior?" or "How can mental states cause physical movements?" might imply that there is a form or structure to the mind, which can be described, either in itself or in its relation to the body or the brain. But instead of these questions we can ask: What are we doing when we do things in our heads? What are the causal relations between mental acts and other mental or overt behaviors? What activities, what forms of mental life, are particular mental behaviors a part of? These investigations require an examination of what and how we do things mentally, what overt behaviors they are related to, and how patterns of mental activities can be understood as constituting forms of mental life. This approach can serve to integrate under a single conceptual framework what might otherwise seem to be disparate problems in philosophy: mental causation, intentionality, actions, speech acts, interpretation of behavior, the will, the relation between overt practices and mental behaviors, etc. I have tried to outline, in this essay, the philosophical framework in which one could begin these philosophical investigations.

I conclude by outlining an empirical study that is also suggested by this approach to mental behavior. It is a study of educational practices, and how they might contribute to differences in the development of the forms of children's mental lives. The study would involve exploring the relation between the overt practices prevalent in a society and the dominant forms of a mental life that may result. I want to stress that this is a speculative sketch of an empirical project — the implications that I draw are tentative and based on a tiny fraction of the empirical research that would be required. I want to illustrate the type of project I have in mind by considering a few overt practices in isolation from the other overt practices to which they are related and with which they co-exist. Thus any speculation
about the effect on the form of a person's mental life is bound to be tentative and over-simplified. However, it should provide an indication of the kinds of implications that might be drawn from a fuller study of the effects of educational and other practices on forms of mental life.

North American educational practices have changed dramatically over the past century. It would be worthwhile considering how different educational practices (as well as the content of the lessons) affect the dominant forms of a child's mental life. In Ontario public schools today, a common practice is to use a 'contract system' for at least some school projects, and in some instances, for the vast majority of a child's daily work. The 'contract', as the name suggests, is modeled on a legal contract. The child and the teacher draw up a contract which outlines what a child will do, when she will complete the work, and in some cases, on what basis it will be evaluated. The child, and sometimes a parent, signs the contract before embarking on the project. One of the motivations for the contract system is the idea that when a child overtly commits herself to a program of work by signing a contract, she is taking control of her work, committing herself to the project and thus giving herself a reason to complete the assignment. In this way the child learns to take responsibility for her own work and develops a sense of pride in completing what she has committed herself to doing. She does this as part of a voluntary quasi-commercial or legal exchange. One can contrast this to a more traditional educational practice of

2 See Knowles (1983) for a detailed discussion of learning contracts, especially in colleges and universities; for learning contracts in elementary schools, see Dennis (1980), Kassens (1984) and Cristiano (1993).

3 "Students accept responsibility for their own learning and gain confidence in their abilities to be self-directed learners." (Cristiano 1993:3) "Students tend to take their work more seriously when they have pledged to agreed-upon assignments and sealed that pledge with their signature." (Kassens 1984:47)

4 The degree to which this is conceptualized as a quasi-legal exchange is captured by the practical advice given to teachers interested in using contracts in their classrooms — "When drawing up the contracts, make them official looking. Reserve a space for students to write in their full name, pledging to complete the contract in the specified amount of time." (Kassens 1984:47)
assignments (which should be familiar to most of us from our own school days) in which a child is told by the teacher what to do, how it will be evaluated, as well as being informed of the immediate consequences of not completing the work (whether failing grades or detentions, etc.)

I am not concerned with the success of either practice in so far as they result in completed assignments, or indeed in 'learning'; both the contract system and the assignment system may produce the same results with respect to how many children complete their work. In contrast, I am concerned with the consequences for the resulting forms of a child's mental life, if the overt behaviors which these educational practices encourage are internalized and become dominant patterns of mental behavior.

If a child internalizes the overt behaviors associated with the contract system, then she may begin to commit herself to future activity by mentally 'signing a contract' with herself, or perform some other similarly situated mental activity. Thus, the child may develop a very individualistic way of being a responsible agent — she takes responsibility and follows through on what she has committed herself to doing in return for rewards she has chosen to pursue. When doing things that she is ordered to do, she may follow through on them by explicitly committing herself to perform those tasks which she is ordered to do — "I'm doing it because I want to, not because you told me to, (and because I see it as an advantageous agreement)." If, on the other hand, a child internalizes the overt behaviors associated with the assignment system, then she may begin to commit herself to future activity by considering what she is expected or required to do, perhaps by mentally telling herself to do something (or imagining another telling her what to do) or performing some other similarly situated mental activity. This child may develop a less individualistic way of being a responsible agent — she acts responsibly by living up to other's expectations or commands. Even when doing things that she has chosen to do
herself, she may do so by telling herself that she must do it or imagining others telling her what to do. Alice does this in Alice in Wonderland:

"Come, there's no use in crying like that!" said Alice to herself rather sharply. "I advise you to leave off this minute!" She generally gave herself very good advice (though she very seldom followed it), and sometimes she scolded herself so severely as to bring tears into her eyes ... (Carroll 1960:21)

A child could transfer these mental behaviors from educational to other contexts. If the child performs these activities repeatedly, they could become part of a dominant form of mental life and have much broader consequences. This kind of transference of behaviors learnt in school into other areas of a child's life is common among school children. When a behavior becomes habit in one type of situation, it is often performed automatically in other, even inappropriate, situations. For instance, children who have learnt to raise their hand when they wish to speak in the classroom will often spontaneously do the same in other situations (such as at the domestic dinner table). A person who has learnt to use a new problem-solving technique at work or at school may apply it to home situations. Indeed, much educational practice in the twentieth century has been aimed towards making education more practical, or more relevant to life outside school, which requires the transference of skills from school to other situations.\footnote{For instance, the introduction to the main curriculum document used in Ontario elementary schools from about 1940 until the mid- to late- sixties instructs teachers, "The school must set up for its pupils an environment in which, through use, they may learn the social techniques, derive attitudes and beliefs, and develop the abilities and skills that social life in a democratic society requires." (Ontario Ministry of Education 1941:6) A more recent document, the Common Curriculum, puts a similar emphasis on the need for socially relevant education. "Because issues in the modern world are interrelated in complex ways, students need to develop a broad understanding of ideas and phenomena and their interrelationships, as well as knowledge and skills that they can apply in a wide variety of subject areas and contexts. The shift in curriculum emphasis to more integrated programming and active, inquiry-oriented learning is intended to assist schools in providing every student with opportunities to develop the independent problem-solving abilities and the knowledge, skills, and values required for effective living in a rapidly changing world." (Ontario Ministry of Education 1993:7)} Children are explicitly encouraged to transfer behaviors related to the contract system into other
areas of their lives. For instance, an American drug resistance program (Drug Awareness and Resistance Education, D.A.R.E.) has children sign contracts to modify their behavior both at home and at school as a way of learning to be more responsible. A daily homework agenda, now common at schools throughout North America, outlines several different techniques for goal setting, planning and documenting achievements, and explicitly encourages children to apply such techniques to all aspects of their lives, school, home and other activities.  

What are the consequences if a child internalizes the contract system, and begins to extend it to other aspects of her life? Such a child could get into the habit of accepting responsibility only for those things to which she has explicitly agreed. Then, it might seem to her that other things to which she has not agreed should remain outside the scope of moral evaluation. When faced with a new rule or expectation, the child must either agree to it (and then later either conform to it or fail to do so by not meeting personal commitments) or ‘opt out’. This can put her in a difficult situation if it is a rule that she is not inclined to agree with (for reasons such as dislike of the rule, inconsistency with other rules or customs, arbitrariness, unfairness, etc.), but which she will be held to regardless (for instance, enforced laws or school rules). In contrast, the child who internalizes the traditional assignment system could get into the habit of accepting responsibility for everything expected or required of her. When faced with a new rule or expectation, the child must accept it (later conforming to it or failing to live up to other’s expectations) or she must reject the authority of those who make the rule. This can put her in a difficult dilemma.

---

6 An agenda used in elementary schools in Canada and the United States encourages students to: "Set goals for school, sports, and your personal life."; "Draw up an action plan for each goal."; "Think of yourself as an achiever. Go for it!"; "Reward yourself when you reach your goals."; "Show your goals to your family members." (Premier School Agendas 1996:5) The agenda provides charts for listing personal goals and tracking achievements. In addition, a 'weekly goals' space is provided throughout the calendar portion of the agenda.
if it is a rule or expectation which is inconsistent with other expectations or rules that she tries to conform to, but comes from a source whose authority she is unwilling to reject, (for instance, her peers, family, work situation).

The forms of a person's mental life might have consequences not only for future behaviors, but also for the types of situations which will cause conflict or difficulty, and for the solutions to these problems. For instance, the first child might have difficulty when faced with authority, and the second when confronted with peer pressure. The solution for the first child might consist in finding reason to accept the rule by finding some way of characterizing the reward so that it is favorable to the child. Thus, the child will have reason to 'sign the contract' or commit herself to future activity — her compliance with the rule in return for some desired reward. The solution for the second child might consist in finding ways to prioritize different sources of external expectations and rules, or in some cases (for instance, when she's in with a bad crowd) to find the strength to reject a source of authority.

A wide variety of overt behaviors taught (explicitly or implicitly) to children may have far-reaching effects on the types of agents that they can become, especially if these behaviors are internalized and become dominant forms of children's mental lives. People do different things in their heads and they do them in different ways. I leave for the future empirical projects such as this one which study the ways in which forms of mental life change, and how these changing forms of mental life affect other behaviors — investigating how a person's overt and mental behaviors develop through a life and within the context of the activities and practices that constitute a society.
REFERENCES


Dennis, Kate. 1980. *Why, How to, and This and That from Howell School.* (Micro-fische.) Box Elder County School District, UT.


Ontario Department of Education. 1955. Programme of Studies for Grades 1 to 6 of the Public and Separate Schools.


TEST TARGET (QA-3)