The Representational Base of Psychological States: Filling a Gap in Dennett's Theory

by

Aaron Nowaczek

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Carleton University

Ottawa, Ontario

Abstract

Daniel Dennett claims that we adopt *the intentional stance* toward other people, by ascribing psychological attitudes to them based on their behaviour (Dennett 1987a; 1987b). Yet we often form attitudes without any observation of behaviour. He says little about the basis of this formation, leaving a gap in his account. I attempt to fill that gap. First, I investigate the intentional stance's assumptions, arguing that Dennett's account works only because we have certain capabilities. Second, I argue that one of our most important capabilities, the capacity to *represent* the world and our psychological life, is the foundation on which we form psychological attitudes. Third, I argue that the attitudes we form are idealized abstractions of this representational content. I fill the gap in Dennett's account by claiming that our psychological attitudes are idealizations of representational content rather than patterns in behaviour, and I do so broadly within Dennett's theory.

Acknowledgements

First, I'd like to thank my thesis supervisor, Professor Andrew Brook, without whose expertise, guidance, and superhuman patience this project would have taken much longer to finish. This project grew out of questions raised during a tutorial given by Professor Brook on the subject of Dennett's views on intentional realism; the project slowly developed over the following year and a half into this final version. The many hours spent with Professor Brook discussing the problems and ideas involved with my thesis were critical to my success in completing this project.

I'd also like to thank the members of my examination committee (in alphabetical order), Professor Mark Macleod and Professor Myrto Mylopoulos, as well as the chair of the philosophy department, Professor Dave Matheson, whose helpful comments, criticism, and discussion during the defense of this project helped refine and clarify various sections of the text.

Lastly, I would like to thank my family; my wife Maryann, whose support was vital for helping me finish; and my fourteen-month-old son Arthur, without whom this project would have been completed eight months ago.

Table of Contents

Introduction.

Dennett's *Intentional systems theory* (IST) ascribes psychological attitudes based on patterns in its subject's behaviour (see Dennett 1978; 1987a; 1998a for clear examples of Dennett's behavioural basis of these attitudes). Yet we often don't ascribe *ourselves* psychological attitudes based on observing our behaviour. Not only are many of our attitudes not self-ascribed, but our attitudes are also frequently kept private even though we could publicly express those private attitudes. How do we form our psychological attitudes? We hardly spend much time observing our behaviour, yet we still form attitudes such as belief and desire. No matter how the attitude is formed, clearly our psychological attitudes aren't based entirely on behaviour. In this project, I'll present an account of psychological attitudes that doesn't require observing behaviour in an attempt to supplement IST. What Dennett says about the psychological attitudes we form consists of a few comments,[1] and there is by no means even a half-developed account in his writings. This is a gap in his theory, and I'll attempt to fill this gap in the following project.

I'll begin the Introduction by acquainting the reader with the notion of *intentionality* as the term will be used in this project, followed by a preamble on the intentional stance, IST, and patterns, before outlining what the problem with IST exactly *is*. Finally, I'll finish the Introduction by presenting the arguments and claims that I offer in the following three chapters to fill the gap in Dennett's theory.

[1] There are some passages in *Consciousness Explained* (1991) that suggest the beginnings of an account, but Dennett doesn't develop this account and in fact he doesn't even seem to consider it necessary. He offers practically nothing about how we form psychological attitudes, other than how those attitudes are ascribed using the intentional stance.

1. Intentionality

The term 'intentionality' means *aboutness*. Psychological attitudes, such as *belief*, *desire, hope*, *intention* (this list is not exhaustive), are all considered to display this aboutness. These attitudes are used to understand, explain, predict, account for, and otherwise describe the behaviour of not only humans but also other animals and sometimes machines. For example, we might ascribe a belief to a person based on that person waiting outside a department store early in the morning (such as, *that person believes the department store will open in five minutes)* to understand (or predict, or explain, or describe, and so on) that person's behaviour. However, psychological attitudes aren't always ascribed *to* a subject; they're just as often expressed *by* a subject. If we were to ask the person waiting outside the department store, that person would most likely say that she[2] believes the store will open in five minutes. An ascription of the belief *the department store will open in five minutes* is based on the subject's behaviour (as well as knowing that the department store will open in five minutes) in Dennett's theory, but what is the subject's expression of that belief based on? This person isn't expressing the belief after an interpretation of her behaviour. For both ascriptions and expressions, the attitude is about the content of the attitude (in this case, when the store will open), but how does either the expressing subject or an ascribing party arrive at their respective uses of the intentional attitude?

Throughout this project, I'll often use the phrase 'psychological attitude' rather

[2] Throughout this project I will use both masculine and feminine pronouns to refer to my particular fictional examples. I will attempt to be gender inclusive with my use of pronouns and attempt to use the masculine and feminine in equal measure.

than ‘intentional attitude’. This is a stylistic choice; psychological attitudes often display intentionality just like intentional attitudes. The purpose of these phrases is to distinguish terms referring to psychological attitudes from non-psychological terms.[3] I am considering only psychological attitudes in this project.

I am analysing Dennett’s theories about intentionality, so I will stick to Dennett’s simple definition of intentionality as aboutness. Psychological attitudes are intentional because these attitudes display this aboutness; the beliefs that are either formed by a person or ascribed to a person are all *about* something. That something can range from events outside the brain, such as whether there is milk in the fridge, to one’s own personality.

The attitude associated with that content is *about* that content. IST and the intentional stance are concomitant theories; the intentional stance is inseparable from IST. Dennett has produced enough articles to fill an entire library about these associated theories, but he has no account of how we form our own psychological attitudes. The next section will introduce Dennett’s intentional stance.

2. The intentional stance

The intentional stance is one of three strategies that can be used to predict or explain a system’s behaviour (see Dennett 1978; 1987a; 2009). Each of these three strategies uses a different set of abstractions to predict and explain behaviour. Explaining or predicting behaviour using the language of physical science is the *physical stance*. For

[3] *The department store will open in five minutes* is not a psychological attitude, but the belief that the store will open in five minutes *is* a psychological attitude. A proposition is psychological only when that proposition is believed, desired, wanted, and so on.

example, examining the electricity flow through the computer hardware implementing a chess-playing program[4] to predict or explain that program's behaviour is an application of the physical stance. This stance alone allows us to "predict the malfunctions of systems"[5] (Dennett 1978, p. 4). A computer that constantly shuts off or won't turn on may require studying the computer's physical constitution to explain *why* the computer is malfunctioning. Perhaps the computer isn't plugged in, or the coffee spilled onto the computer by a colleague is causing the computer to malfunction. The physical stance comes with some costs; explanations using this stance are usually very complex. Specific tools and knowledge are required to make sensible physical stance explanations.

Predicting the computer's behaviour by considering that system's physical constitution will take a long time and require specific knowledge about the laws and mechanisms involved with the computer implementing the chess program. Another way to predict this program's behaviour is to use the *design stance*; when you attempt to predict the behaviour of a system by looking at that system's *design*, you're adopting the design stance. A chess program is designed to play a good game of chess; we can predict or explain the program's behaviour by assuming that the program is functioning as designed. This strategy assumes that not only has the system been *designed*, but will also function according to how the system has been designed to function (Dennett 2009, p. 3). The design stance yields swifter predictions than the physical stance because the design stance assumes only that the system in question will function according to its design. The physical constitution of that system is ignored by the design stance. The physical stance

[4] A chess program is Dennett's earliest example of a system whose behaviour can be predicted or explained using all three stances. See Dennett 1978, p. 4.
[5] Unless, of course, the system is designed to malfunction; in this case malfunction would be part of the system's design (Dennett 1978, p. 4)

can be costly because it requires knowledge about the physical states of the system, along with knowledge concerning the physical laws and mechanisms operating within the system. The design stance, however, takes the physical stance for granted and adopts two critical assumptions: first, that the system *is* designed, and second, that the system will function *as it was designed to function*. A chess program was designed by engineers to play chess, and by playing a game of chess against the program, one is assuming that the program will function as designed. The cost of the design stance is that it can explain a system's behaviour by looking *only* at a system's design. The strategy ignores everything about how that particular design is realized in a physical system. A chess program is implemented on computer hardware, but functioning computer hardware is taken for granted when the design stance is applied to the chess program. The design stance is more efficient than the physical stance because the design stance takes some features for granted.

In addition to the physical and design stance, Dennett argues that we can also use the *intentional stance*. Whereas the physical stance interprets a system's physical constitution and the design stance interprets a system's design, the intentional stance interprets behaviour and ascribes psychological attitudes to a system. In the case of the chess-playing computer program, when neither the code nor the rules and tricks programmed into the chess program are immediately available, we can switch to the intentional stance "by treating the machine rather like an intelligent human opponent" (Dennett 1978, p. 5). This strategy is swifter and more efficient than either the physical or design stances because not only is the system's design and physical constitution taken for granted, but also the system's rationality. This rationality assumption allows one adopting

the intentional stance to ascribe *beliefs*, *desires*, and other psychological attitudes to the system for predicting and explaining the system's behaviour.

The intentional stance requires that the system is both *rational* and *designed*. A *rational* system is a system that follows its desires using its beliefs while minding the constraints imposed on the system. *X* wants the light turns off, so *x* gets up and turns off the light switch. There's nothing constraining *x* from satisfying the desire (such as a power outage or paralysis preventing *x* from turning off the light), and *x*'s action is a *rational* action because she's using her beliefs to satisfy her desires. If *x* is rational, then *x* will behave as if she has certain beliefs and desires and pursue those desires using her beliefs; *x*'s behaviour can be either explained or predicted by ascribing to her the beliefs and desires that she ought to have.

The system's *design* (evolutionary design in the case of humans and other animals, with natural selection as the designer) is the second assumption of the intentional stance. The subject of this stance is designed to have the beliefs and desires that it *ought* to have because that subject's evolutionary history would have granted particular capabilities to that subject. Two basic examples of these capabilities would be the need for food and a capacity to gather information about the world; we're biologically equipped to navigate the world and find sustenance.

The intentional stance is used to predict and explain the behaviour of rational, designed systems. These two assumptions accompanying the intentional stance come with some trade-offs, however. One trade-off is that psychological attitudes are supposed to be causes of behaviour, except that, as causes, they are "surely as unspecific and unhelpful as a causal explanation can get" (Dennett 1987b, p. 57). Explanations using this

stance are quick and cheap; intentional stance explanations require different kinds of background information than either the physical or the design stances.

Another trade-off is the neutrality embraced by the intentional stance toward the implementation of its subject's design. For instance, the stance can be adopted toward a chess-playing computer program or a human as long as ascribing beliefs and desires will predict the system's behaviour. Each of these systems was designed for different tasks and implements those beliefs in completely different ways. Humans are biological; programs are not. The intentional stance ignores this implementation difference. In other words, the implementation of the design is taken for granted as a consequence of the assumption of design.

If any of the strategies break down, then a different stance can be used to explain behaviour. When a computer program doesn't function as it was designed to function, shifting to the physical stance and examining the hardware would be a good place to start for predicting and explaining the program's behaviour. Similarly, if the intentional stance doesn't work or isn't immediately available, one might attempt to use the design stance to predict the system's behaviour. Dennett uses the example of severely counter-suggestive people to illustrate the movement from using intentional explanations to design stance explanations (Dennett 1978, p. 10). If a subject isn't behaving rationally, then the intentional stance won't work, and the design or the physical stance must be used to explain the counter-suggestive subject's behaviour.

One can shift from stance to stance even if none of the stances has broken down. The only real limitation on shifting from stance to stance is the availability of the stance to which one wants to move. For example, if I'm predicting the behaviour of a chess-

playing computer program and I don't understand anything about computer hardware, then the physical stance won't be available to me. A chess program's behaviour can be explained in different ways, either by an opponent using the intentional stance, the program's designer using the design stance, or the computer repairman using the physical stance.

Each stance is distinct. The three stances include different assumptions and corresponding trade-offs, and this distinctiveness is defined, in part, by those assumptions. For example, the assumption of rationality is unique to the intentional stance; the psychological attitudes used by the intentional stance are dependent on rationality. Neither of the other two stances assumes rationality in their predictions. Rationality is visible only when psychological attitudes are used to explain or predict behaviour. The design stance sees design; the intentional stance sees rationality.

The intentional stance ascribes intentional, psychological attitudes to systems. These ascriptions are not based on a detection of discrete states corresponding to those attitudes in the system's brain, or "finding something inside the believer's head" (Dennett 1987a, p. 14). Instead, the intentional stance ascribes those attitudes on the basis of a system's behaviour. Behaviour is the most significant aspect of the intentional stance. The next section will continue the discussion of the intentional stance and IST, but shifts to the intentional stance's reliance on *patterns* in behaviour for the ascription of psychological attitudes.

3. Patterns in the world

Knowledge about a system's physical states is required for the physical stance to

be effective, and for the design stance, an assumption that system is not only designed, but also designed well is required to make predictions and explanations. The intentional stance treats its subject as a rational agent and ascribes attitudes to that agent based on *patterns* in the agent's behaviour. An agent's behaviour contains a *pattern* when there is an explanation for that behaviour that's faster and more efficient than an explanation relying on the whole record of behaviour (Dennett 1998b). For example, we could produce a written record of all the minute, individual behaviours that an agent undergoes while performing a simple task, such as folding laundry. One way to explain that agent's behaviour would be to rely on that record, but that method would be both immense and arduous; folding the laundry requires a huge number of small distinct behaviours and motions. Alternatively, we could interpret the pattern in that behavioural record as simply *wants to finish the laundry tonight*. According to Dennett, both the comprehensive, written record and the pattern in that record can provide an explanation for the agent's behaviour, but using a psychological attitude to explain the pattern in an agent's behaviour is a swifter and more efficient way to produce an explanation than the record.

These attitudes are *idealizations* of patterns in behaviour. A pattern is *idealized* by abstracting away from the messy details and using an attitude to explain the pattern in behaviour. For instance, we could ascribe a *desire to be an MP* to a person running for political office based on that person's behaviour. Rather than provide a list of all the various ways the person behaves to explain that behaviour, we can abstract away from the minute details of behaviour and idealize his behaviour by ascribing the *desire to be an MP*.

If that politician's behaviour can be predicted successfully using an attitude, then

the politician is said to have that particular attitude. Dennett makes this point another way: "*what it means* to say that someone believes that *p*, is that that person is disposed to behave in certain ways under certain conditions" (Dennett 1987b, p. 50). Idealizing the pattern in the politician's behaviour produces quick, simple, and generalized explanations of behaviour (the fine precision of the other two stances isn't required in this case).

G*eneralized* explanations mean explanations that are *abstract*. Attitudes are *abstract* because first, the same attitude can be ascribed to many different behaviours (a belief can be ascribed to many different patterns in behaviour), and second, an attitude ascription is based on a pattern in the subject's *behaviour* rather than knowledge of a particular brain state or mechanism happening in the subject. The pattern in behaviour is idealized using an abstract psychological attitude. Since attitudes are abstract, they can be used to explain many different behaviours: for instance, person A is checking the weather online under a shelter, and person B is in the vicinity walking outside. Each individual could be ascribed a belief about the weather, even though the basis of that ascription is different. Using the intentional stance, abstract attitudes are ascribed based on patterns in behaviour and the particular attitude will depend on how the subject's behaviour is idealized. In Dennett's view, idealizing a pattern in behaviour using an abstract psychological attitude results in a fast and efficient way of explaining behaviour.

Systems whose behaviour can be explained or predicted this way—using the intentional stance—are called *intentional systems*. These intentional systems are predicted and explained as *systems*: "individual beliefs and desires are not attributable in isolation, independently of other belief and desire attributions" (Dennett 1987b, p. 58). Systems aren't ascribed beliefs and desires in isolation from other beliefs and desires, and

any successful ascription will have been mindful of that system's other beliefs and desires for the ascription. Intentional systems are properly *systems* of psychological attitudes. At any time, a system can be attributed different, and possibly interconnected, psychological attitudes.

4. What is the problem?

Dennett is not *entirely* silent about the formation of our psychological attitudes, but we have to do a little interpreting and digging to find where he breaks his silence on this subject. For instance, in *Consciousness Explained* (1991), he claims that various brain processes are *probed*, which produces different narratives (p. 135). Here is a clear way of understanding a probe from a different article:

> A probe is a new stimulus that draws attention (resources) to a particular area…thereby promoting the influence…of whatever is occurring there and rendering it reportable and recollectable. (Akins & Dennett 2008)

Basically, brain content is selected by a probe, and that selected content forms a *narrative*: a version of events that plays a role in regulating behaviour (Dennett 1991, p. 254). Generally, Dennett's idea is that the brain selects content using a probe—giving that content more resources—and probed content forms narratives that may find their way into a psychological attitude.

Dennett may not be *entirely* silent, but what he says is not overly helpful for explaining how psychological attitudes are formed. At no point does he spell out how attitudes are formed, except using vague concepts that you and I must interpret, develop, and apply to an area (psychological explanation) that Dennett isn't obviously talking

about. The vague and brief comments about *narratives* and *reports* doesn't give us a great place to figure out how attitudes are formed. Psychological attitudes are the subject of IST, not *Consciousness Explained*, and IST focuses almost exclusively on behaviour as the basis for attitude ascriptions. Other than a few comments here and there, Dennett provide no basis for the psychological attitudes that we form, attitudes that make up our psychological life.

Behaviour is critical for the intentional stance and IST. In the case of humans, *behaviour* may take the form of speech (as in what an individual says) or in the form of action (as in what an individual does). The *desire to do laundry tonight* may be ascribed to an individual based on what he's doing (collecting laundry into a hamper, for example) or what he says (he informs you that he needs his nice clothes tomorrow, but they're dirty), or even using both what he does and says. In both cases, behaviour is the critical factor for ascribing a psychological attitude.

One problem is that behaviour is not the only basis for our psychological attitudes. Dennett claims that the intentional stance is unavoidably adopted with regards to both oneself and other people (Dennett 1987a, p. 27; 2009, p. 1) but he focuses his discussions exclusively on the adoption of this stance toward other people. In the case above, on what basis does the man express the desire to have his clothes tomorrow? That basis may consist of facts about his plans for the next day, but the key point I want to emphasize is that he isn't basing his desire on an observation of his behaviour.

How does he form the desire that he expresses? An *expression*, as I will use the term, refers to the attitudes we use to communicate ourselves publicly. The primary difference between *expressions* and *ascriptions* is that while an individual expresses his

attitudes, a subject is *ascribed* an attitude by either another person or himself.

That being said, ascriptions are also made *by* someone. That would, using my definition, mean that an ascription is a kind of expression. The salient difference is that an ascription is also made *to* a subject, whereas an expression is made *by* a subject. An expression can also be reported *to* someone (such as when you tell somebody one of your beliefs), but expressions are distinct because they aren't solely based on behaviour. The difference between an attitude that's ascribed to someone, versus an attitude that's formed (and subsequently expressed, or self-ascribed, or reported, or articulated privately) by someone, is one of the topics of this project.

Here's a simple, yet fascinating example of a belief expression that isn't based on observing one's behaviour. In "Intentional Systems in Cognitive Ethnology" (1987c), Dennett muses on whether or not vervet monkeys "really" communicate (pp. 238–239). Vervet monkeys appear to have rudimentary communicative capacities, with the ability to make specific vocalizations in response to different predators—such as eagles, snakes, and cougars—which are believed by researchers to express the presence of the predator to the rest of his troop. A vervet monkey giving the vocalization for eagle, for example, to alert his troop of the approaching predator is displaying what's on his mind. In other words, the monkey sees or hears the eagle, causing him to believe that there's an eagle in the sky, and he transmits what he sees—his belief—via a warning vocalization to the rest of his troop. The monkey's ability to call in response to a predator is a display of not only information that he has about the world, but also his cognitive capacities. The monkey *sees* the approaching eagle, implying that he has some form of mental representations involving the eagle, and he can vocalize this fact, which suggests the monkey has certain

cognitive capacities. The fact that the monkey is making an eagle vocalization means that his brain is capable of responding to the environment with an appropriate vocalization. In this case, the monkeys are *expressing* an attitude—a belief about a particular predator—and the intentional stance uses this expression in an attempt to predict that monkey's behaviour. The monkey is not inferring from his own behaviour that there's an eagle in the sky, he *saw* the eagle and expressed his belief. For this expression to be possible, the monkey must be able to express attitudes without any inferences about his own behaviour.

Vervet monkeys have the capacity to express beliefs that aren't inferences about their own behaviour, and so do humans. The intentional stance can use those expressions to explain the monkey's behaviour, but the stance can't explain how the monkey can express his belief about the eagle. The monkey is clearly not basing his vocalization on his own behaviour; he's basing his expression on his visual experience of the eagle in the sky. In Dennett's view, patterns in behaviour are the most common basis for psychological explanation.

Not all expressions have to be verbal; gestures are frequently used to express an attitude. Sometimes, the only way to express anger or frustration to a fellow driver making poor decisions on the road is to gesture with your middle finger. Note that in this example my attitude isn't based on my behaviour. I'm not ascribing the attitude based on my behaviour using the gesture; I'm expressing my attitude using a gesture.

What if I decided not make my attitude public? Some attitudes are kept private, and I will call these private attitudes *private articulations*. I choose the term 'private articulations' for these private attitudes simply to distinguish these private articulations

from the more common *public* expressions that I've discussed so far. To *articulate* something is to make that something clear and distinct, and both private articulations and public expressions often involve making something clear or distinct using words. In this project I will use *private articulations* to refer to the attitudes we keep to ourselves; expressions are made public but unlike private articulations we can express attitudes non-verbally. The not-so-friendly gesture from the previous paragraph is an example of a non-verbal expression. Unless otherwise stated, when I use the term 'expression' in this project I do not discriminate between verbal and non-verbal forms of expression.

Most often, patterns in behaviour serve as the basis of psychological attitude ascription in Dennett's IST even though some attitudes can be formed *without* any input from behaviour. IST and the intentional stance offer an analysis of psychological attitudes that relies on behaviour, so whether an attitude is expressed, ascribed, or privately articulated, the basis must be found by looking mostly at behaviour. The source of the beliefs and desires we express or privately articulate requires a different analysis than the beliefs and desires other people might say we have.

We don't exactly observe our behaviour with the same ease as other people. Usually, we're too busy with actually *behaving* to pay attention to our behaviour. If behaviour is held to be the primary basis of our attitudes, as Dennett often claims, then what about the attitudes we express not on the basis of behaviour?

5. This project

I am convinced that the intentional stance does a good job of explaining how and why psychological attitudes are ascribed to other people (and animals, and machines).

We don't need direct access to what's going on inside our subject's head to say that the subject believes or desires something. We access what's going on inside the subject's head through their behaviour. But there is a difference between my relation to my own attitudes and the relation that I have to other people's. This difference can be summed up as: we don't express other people's attitudes; we ascribe attitudes to other people.[6] The intentional stance can't accommodate this difference, because of its reliance on behaviour. I don't form my psychological attitudes on the basis of my behaviour, so the attitudes I express, ascribe, or privately articulate must be based on something other than an observation of behaviour.

This project will attempt to fill a gap in Dennett's theories regarding how we form our psychological attitudes. I will analyze the framework of IST to answer this question. Here's a highly condensed version of what this project aims to achieve. The intentional stance requires certain capabilities for its predictions. These capabilities are displayed in behaviour, and represent what's going on inside the brain. By investigating these capabilities, we can see that one of these capabilities involves the brain's capacity to represent the world and oneself. I will claim that these mental representations form the basis for our psychological attitudes. We access those representations through the attitudes we form, and those attitudes are idealizations of brain content that's been abstracted away from the rest of the representational content. The representations that have been abstracted away from all of the messy details are idealized using a psychological attitude, which forms the content of our psychological life.

I am broadly adopting Dennett's ideas for my account. My account aims to

[6] This simple and helpful way of stating the difference I'm referring to was suggested to me by Professor Andrew Brook.

supplement IST, and seeks to provide an explanation of the source of psychological attitudes that we ourselves form, publicly express, and also ascribe. Here is a slightly more detailed overview of the project that follows. In Chapter I, I will argue that the intentional stance requires that its subject displays its evolutionarily derived capabilities in behaviour. A system that can have its behaviour predicted or explained using the intentional stance must have certain capacities. For instance, very generally, an intentional system must be able to respond appropriately to the world. That is, the system must have a capacity for gathering information related to its environment to survive. I will use the term 'information' very generally; sensory information and information contained in memories are both examples of capacities that a system must have for that system to be an *intentional system*. If an intentional system is both rational and evolutionarily designed, then that system must be evolutionarily designed to have capabilities for realizing or implementing rational behaviour.

Chapter II provides a more in-depth analysis of an intentional system's capabilities which allow for the expression of psychological attitudes: attitudes can be self-ascribed, expressed, articulated, and so on, by oneself based on the brain's representational content. *Representational content* replaces *behaviour* as the basis for the psychological attitudes we form in the view that I'm presenting. These mental representations refer to the various stimuli represented by the brain. Visual experience, memories, other attitudes, feelings, blood pressure, hydration level, etc., are all examples of information that the brain can represent.

For example, *x* sees a tree. That tree is represented in his brain. Here's another example: *x* is jealous of *y*. That jealousy of *y* is represented, somehow, in *x*'s brain.

Exactly *how* representations are implemented isn't the topic of this project. I will be focusing on the *content* of those representations. It doesn't matter if the content's source is (for example) something visual (such as the tree) or some bodily function (thirst); it only matters that content *is* represented.

Psychological attitudes are *about* something; mental representations are *also* about something. These states share *aboutness*. Even though mental representations and intentionality share aboutness, they don't share rationality. Psychological attitudes are a subcategory of representational content generally, a subcategory that arises from the rationality conditions applied to attitudes. Successful attitude ascriptions require rationality, but rationality doesn't have to apply to mental representations. There doesn't have to be anything rational about how the brain represents the world: the brain simply represents the world as it was designed to represent the world. Seeing a tree, and thus having some mental representation about a tree, has no direct connection to rationality.

Chapter III investigates the relationship between representational content and our psychological attitudes more closely. The view I present is somewhat similar to Dennett's, except with one major difference. Instead of abstracting and idealizing attitudes based on the behaviour of another person (as Dennett claims), I will claim that the brain goes through this process of abstraction and idealization for its *own* content. My view is that psychological attitudes have a representational basis, and attitudes are abstracted from other content and subsequently idealized as a particular attitude.

Furthermore, how that idealization relates to the rest of the system's mental life will dictate which attitude is used in the idealization. In other words, representational content is idealized using a particular attitude depending on the context of that

idealization. This is a holistic approach to psychological attitudes, and the approach is largely based on the idealized content's *direction of fit* (a concept which will be discussed in Chapter III, section 3.7). A belief expression, for example, is also an expression of the relationship between the expresser and the belief's content. When a person expresses a belief *that the Bigfoot lives in Saskatchewan*, that expression is based not only on representational content but also the role that content plays in the system's mind. That person is expressing information that she possesses (in this case, where the Bigfoot lives) in the form of a belief.

The arguments that I will provide in this project present a major addition to IST. This supplement attempts to provide an account of how we form the attitudes that compose our psychological life, while attempting to fill the gap in Dennett's theories.

Chapter I.

Intentional Systems and their Capabilities

1.1. Introduction

There are certain features of an intentional system that can provide a basis for forming psychological attitudes without relying on behaviour. An intentional system's cognitive capabilities, for example, are critical for intentional stance ascriptions because of the way those capabilities are connected to behaviour. By investigating these capabilities, I will set the stage for my account of the basis of our psychological attitudes.

This chapter will argue that intentional stance ascriptions require a system to have certain capabilities.[7] These capabilities are displayed in the behaviour interpreted by the intentional stance. Moreover, this display has some design implications for an intentional system. Some of these design implications form the basis for my account of psychological attitudes, and I will begin my investigation by considering these implications. The intentional stance bases psychological attitude ascription mostly on observing behaviour, but many of the psychological attitudes we form are *not* based on observing our behaviour and thus aren't the result of adopting Dennett's version of the intentional stance.

Investigating the capabilities required by the intentional stance is the first step in the account of psychological attitudes provided in this project. Since I'm attempting to provide a supplement to Dennett's IST, I aim to remain broadly within the framework of

[7] I will use the term 'capabilities' to refer to an intentional system's capabilities *in general*; the term 'capacity' or 'capacities' refers to *specific* capabilities. Here's an example: an intentional system has *capabilities* involved with responding to its environment, such as the *capacity* for gathering visual or auditory information.

Dennett's theory. Drawing consequences out of Dennett's version of intentional systems will all me to develop an account of how we form our own psychological attitudes while remaining roughly within the framework of IST. My account constitutes a major supplement to IST and the intentional stance, but the seeds of that supplement are contained in Dennett's theories. By investigating the design requirements of an intentional system, I will argue in subsequent chapters that some capabilities are invisible to the intentional stance. This chapter will argue that the intentional stance—and IST—depends on its subject (chiefly, us) having particular capabilities, ones that allow for successful predictions and explanations of the subject's (our) behaviour. In fact, both the intentional stance and IST *require* that its subject display these capabilities in their behaviour.

This chapter will proceed as follows. In section 1.2 and 1.3, I will argue that brains are connected to psychological attitude ascriptions through behaviour. The capabilities required for any particular psychological attitude ascription are evident in a system's behaviour that the ascriptions are based on. The behaviour of humans and other animals is controlled by the animal's brain, and a brain's capabilities will govern the patterns in that animal's behaviour.

Next, in section 1.4, I will claim that behaviour is a *display* of the brain's content and capabilities. I am using the term *display* in the ordinary conversational sense; brain contents are *displayed* in behaviour because patterns in behaviour show something about what's happening in the brain, as well as what the brain is capable of doing. Another reason to think of behaviour as a display rests on how brains respond to the world using various behaviours. These various behaviours may take the form of (for example) verbal

communication, gestures, or actions. Additionally, some brain responses are unconscious, and many of those responses are *forever* unconscious (such as hormone production). Behaviour is accessible to other people (you can *watch* someone else behave, or *hear* someone else speak), whether that behaviour is caused by conscious or unconscious processes, and that behaviour can be interpreted using the intentional stance. The system's capabilities (such as having sense organs and responding to sensory stimuli) allow the system to respond to the world and also display those capacities using behaviour.

In the final section, 1.5, I will argue that intentional systems must have particular design features. The content of some psychological ascriptions strongly suggests that there must be something particular going on in the system's brain. For instance, singing a song requires that you've heard the song, and both IST and the intentional stance require that the brain operates a certain way (generally, operating in a way allowing for verbal outputs that reference aural inputs). This operation will be the basis for my account of psychological attitudes in later chapters.

This chapter shows how investigating the capabilities of an intentional system can set the stage for an account of psychological attitudes that aren't formed with behaviour as their basis. Once we realize that a system can display its brain contents, looking at the capacities related to those contents will provide insight into our attitudes. An intentional system's behaviour is a consequence of that system's capabilities and the ascription of those capabilities, as I will argue in this chapter, is derived from IST's analysis of intentionality.

1.2. Brains and their capacity to behave

In this section, I will argue that ascriptions are connected to the brain because ascriptions are based on a brain's capabilities, capabilities that can cause patterns in behaviour. Investigating this connection will provide insight into the capabilities inherited by an intentional system from that system's evolutionary heritage because ascriptions are based on behaviours caused by the brain's capabilities. The intentional stance uses these evolutionarily designed capabilities for its predictions and explanations of behaviour.

This connection between patterns in behaviour and the brain is an important if obvious point that needs to be recognized to its fullest extent. The way that people behave is a direct result of brain activity. This is an uncontroversial point, but I don't think that Dennett adequately examines the consequences of this causal relationship and its impact on the intentional stance. He's typically focused on the important, and very interesting topic of the *reduction* of psychological attitudes: beliefs, desires, fears, etc., don't reduce to concrete, individualized types of brain states in IST (see Dennett 1978; 1987a; 1998a). Dennett often tells us how beliefs and desires are *not* related to the brain, but he never really says anything about how beliefs and desire *are* related to the brain. What he does say is that psychological attitudes are ascribed to systems based on patterns in that system's behaviour.

For example, if a person ducks a baseball thrown in his direction, then we can explain that particular behaviour by ascribing a belief to the person about the ball's trajectory, or a desire about not wanting to get hit by the ball. The way the person behaves is interpreted and ascribed an attitude using the intentional stance. I agree with

Dennett that attitudes can be ascribed in this way (based on patterns in behaviour), but more can be said about the person in this example to explain why he ducks the baseball. In order for a person to *duck* a baseball—in other words, in order for that person to have a pattern of behaviour explainable by ascribing a belief or a desire based on an approaching baseball—that person *must* have certain capabilities *enabling* him to duck the baseball. That person couldn't duck if he was blind or immobile, for example.

Dennett wouldn't disagree with this fact. *Of course* intentional systems have capabilities (such as those capabilities involved with ducking a baseball); intentional systems are evolutionarily designed systems. These capabilities have survival value, "however [they happen] to be realized as a result of mutation" (Dennett 1987b, p. 59). It's simply good design to respond to threats in one's environment, for example. We don't have to know *how* the system's capacity to avoid harm is implemented, knowing that the system has the capacity is enough. Not only is an intentional system designed to respond to its environment, the system is designed to respond appropriately to its environment: an intentional system *is* a rational and evolutionarily designed system, so it should be able to respond appropriately to many different circumstances in its everyday life. The connection between that person's brain and the patterns in his behaviour warranting the belief ascription is the issue in this section.

The importance of examining an intentional system's capabilities for this project is that some of these capabilities allow an intentional system to form psychological attitudes. This examination is, in my view, the best place to begin my account of the psychological attitudes. I introduced how Dennett conceives of *patterns in behaviour* in section 3 (pp. 9-11) of the Introduction to this project, and in the following section, I will

begin by expanding on that introduction to show how patterns in behaviour are connected to the brain.

Let's start by looking further into Dennett's concept of *patterns in behaviour*. Patterns are fundamental to the intentional stance, and Dennett provided the first vivid example of this relationship in "True Believers". In reply to Robert Nozick about the importance of the intentional stance, Dennett re-purposes Nozick's example of super-intelligent Martians (who have mastered the physical stance) to show the intentional stance's necessity:

> Take a particular instance in which the Martians observe a stock broker deciding to place an order for 500 shares of General Motors. They predict the exact motions of his fingers as he dials the phone and the exact vibrations of his vocal cords as he intones his order. But if the Martians do not see that indefinitely many *different* patterns of finger motions and vocal cord vibrations–even the motions of indefinitely many different individuals–could have been substituted for the actual particulars without perturbing the subsequent operation of the market, then they have failed to see a real pattern in the world they are observing. (1987a, pp. 25-26)

Without the intentional stance, the Martians wouldn't recognize the many different ways of realizing the same outcome, the same *desire to buy GM stocks*. According to Dennett, these Martians cannot make intentional explanations about the stockbroker's behaviour. Detecting those patterns in behaviour requires the intentional stance, which the Martians don't have. There *are* patterns in the stockbroker's behaviour, but the Martian's physical stance cannot, by definition, detect those patterns.

It wasn't until 1991 that Dennett published a thorough attempt to explain how there *really are* patterns in human affairs. Prior to "Real Patterns" (1998a)[8], Dennett had insisted that there are *real patterns* in human affairs:

> There *are* patterns in human affairs that impose themselves, not quite inexorably but with great vigor, absorbing physical perturbations and variations that might as well be considered random; these are the patterns that we characterize in terms of the beliefs, desires, and intentions of rational agents. (Dennett 1987a, p. 27)

Here's another example from later in the same text:

> intentional stance description yields an objective, real pattern in the world—the pattern our imaginary Martians missed. (Dennett 1987a, p. 34)

These quotes are only two examples of Dennett's view, and there are more. Both examples sufficiently illustrate his long-held conviction that there are real patterns in the behaviour of rational systems, patterns that are exploited by the intentional stance to predict or explain behaviour.

Interestingly, despite the importance of "Real Patterns" to Dennett's view, he never *directly* connects the discussion in that paper to human behaviour. Instead, "Real Patterns" is a defense of the idea that a pattern in a dataset provides a more efficient description of the data than a description including the pattern along with the noise accompanying that pattern (Dennett 1998a, p. 103). *Noise*, in this context, refers to data that isn't part of the pattern but exists alongside the pattern in the dataset.[9] Even though

[8] "Real Patterns" was first published in 1991 and later reprinted in *Brainchildren* (1998).
[9] I'll make this clearer. Say I threw a handful of pennies on the floor and some of them landed in such a way to create a smiley face. In this case, the *smiley face* is a pattern in the handful of pennies now on the floor, and the pennies that *aren't* part of the smiley face constitute the noise. The handful of pennies is the dataset; the smiley face is the pattern; the pennies that aren't part of the smiley face constitute the noise.

Dennett doesn't *directly* make the connection between patterns in data and patterns in behaviour, it's not difficult to see how patterns in data are analogous to patterns in behaviour. For example, if we accept Dennett's conviction that there *really are patterns* in human behaviour, then it's a short step to see that it's faster and more efficient to ascribe a *desire to buy GM stocks* to the pattern in the stockbroker's behaviour than it is to provide a physical stance explanation of *all* the stockbroker's behaviour.

Here is a simple, if somewhat archaic, example of a real pattern and a pattern detector. A fax machine is a mechanical example of a pattern detector. When a fax machine scans a document, say a document containing English prose, the machine creates a bitmap of the document based on the locations of the light and dark sections on the document. The machine maps these locations onto a grid, where each tiny square is either light or dark depending on whether the corresponding area of the document contains text or not. This bitmap, representing the document as a coordinate grid containing white and black squares, is transmitted through the telephone lines to another fax machine, which reinterprets the bitmap back into the original document. The representation created by the fax machine is the bitmap, and even though the bitmap can be used to reconstruct the original document without another fax machine (as long as you have a set of instructions about how to translate the representation into the original document), this representation is a far more cumbersome and less efficient way of transmitting the information contained in the original document than a simple reading of the original document's English prose. The information transmitted by the fax machine contains a real pattern—in this case, the English prose—that's a more efficient way (to readers of English) to interpret that information than a description of the bitmap representation created by the

fax machine.

If we didn't have a fax receiver (or an instruction manual), or if we couldn't adopt the intentional stance similar to the Martians, then we wouldn't be able to read the prose contained in the bitmap or be able to see the desire in the stockbroker's behaviour. Even if all the fax receivers in the world magically disappeared, along with the collective knowledge that fax machines even exist, the information transmitted by a fax machine would still in principle contain a real pattern. The pattern in both the bitmap and the list would be indiscernible without a fax receiver or the intentional stance. In the same way, an intentional stance adopter may fail to recognize a real pattern in behaviour because that pattern is indiscernible to the adopter.

Just because a pattern is indiscernible, however, doesn't mean that pattern isn't there. For example, "other creatures with different sense organs, or different interests, might readily perceive patterns that were imperceptible to us all along" (Dennett 1998a, p. 103). A creature that could see light in the ultraviolet spectrum might be able to detect features imperceptible to us. For instance, without a device allowing one to see the UV spectrum, we would have no idea that bees use UV light to find flowers.

UV light causes patterns in the bee's behaviour, and we wouldn't be able to know that a particular pattern in the bee's behaviour is a result of seeing UV light unless we had a tool or machine for detecting UV. We might still be able to detect a pattern in some bee's behaviour because she's[10] responding to the UV light, but without tools to detect the UV spectrum, we wouldn't know the bee was responding to a different light source.

If we could detect the UV light, then we'd be able to explain new patterns. A

[10] Worker bees are female.

bee's detection of UV light would be known by us, and we could use that behaviour—along with our new knowledge of the bee's capacity to discern light in the UV spectrum—to make further predictions. The bee's brain responds to UV light, and interpreting the bee's behaviour based on this capacity is also an interpretation of the bee's brain content. In other words, the bee has the capacity (gifted by its evolutionary heritage) to see UV light. The formerly indiscernible patterns in the bee's behaviour, which were caused by UV light, can give an intentional stance adopter information about the contents of the bee's brain. The bee can *see* UV light, a light source indiscernible to humans without instrumentation, and that particular visual capacity is a cause of the bee's behaviour.

The bee uses the UV spectrum to find nectar. Unless we knew that bees could see UV light, we wouldn't know that they use the UV spectrum to find nectar. Ascribing an attitude based on the bee's behavioural patterns, in this case, is also ascribing an attitude based on the bee's ability to both *detect and respond* to the UV spectrum. The pattern is an effect of this ability. The only reason that we could say a belief or a desire motivated the bee, in this case, is if that bee had certain information, motivations, or goals (involving UV light reflected from certain flowers) resulting from its response to the UV spectrum.

The bee example provides an illustration of patterns regularly indiscernible to humans. Just because we can't detect a pattern somewhere doesn't mean a pattern isn't there. We can often see the effects of an indiscernible pattern before we learn of the pattern's existence. Knowing that the bee can see UV light will allow for an ascription based on seeing UV light, even though the bee's behaviour doesn't change based on our

pattern detection ability. If we ascribe a belief or a desire based on the bee's ability to see the UV spectrum, then the particular attitude is connected to that ability. The bee's ability causes the patterns in her behaviour, and thus the ability is assumed when that particular ascription (a belief that *there is suitable nectar in this flower because of UV markings*) is made based on a pattern. Detecting the pattern allows us to ascribe this belief based on the bee's ability. This belief only makes sense because the bee can detect UV light. In this case, a bee's ability to see the UV spectrum is necessary for an ascription of that particular belief. This ability persists regardless of whether or not an observer could detect it.

Ascribing the belief *there is nectar in this flower* to a bee based on the bee's behaviour suggests that she can see the UV markings, which connects this ability (detecting UV light) to an attitude (the belief about the flower). Somehow, the brain implements the bee's capacity to see UV light. Presently, the exact manner of implementation is not important; what is important is that this implementation exists in principle. A bee's brain is connected to the attitude because the way that the bee's brain responds to the environment will dictate which attitude is appropriate for ascription. The bee can see UV light, and thus have beliefs based on UV light. This is an example of how specific behaviour is a result of the brain's capabilities, and the intentional stance interprets these capabilities using behaviour.

This bee example also shows how a system's capabilities are a necessary factor for ascribing psychological attitudes based on the system's behaviour. Seeing UV light causes the bee to behave in a certain way (finding nectar using UV indicators on flowers), and our ascriptions based on the bee's behaviour are a direct result of this particular

capacity. It doesn't matter if we know that the bee can see UV light or not. We don't even have to know that the bee can see *anything* to ascribe a belief based on the bee's hunting for nectar, all that's required is that the bee has *some* capacity to gather information about the world. As long as the bee is behaving *as if* she's hunting for nectar, the intentional stance can ascribe a belief based on the bee's behaviour, regardless of whether we know how or why the bee is hunting for nectar.

Unlike the bee example, it's not always easy to see how the brain connects to psychological attitudes. Humans are more complicated beasts than bees. The bee example illustrates how *what a system is doing* is directly connected to *what the system can do*, i.e., its capabilities. Another way to make this point is to say that a system's capabilities will determine the kinds of content that can be ascribed using the intentional stance. We wouldn't ascribe a belief to a person based on an ability to see the UV spectrum unless that person was using a tool to detect UV light. Put simply; ascriptions are made to people based on the things people can do.

There are an uncountable number of attitudes that can be ascribed to any person at any time based on what the person can do. For instance, someone could ascribe a belief to me that *The Intentional Stance* is on my shelf right now, or that the window blind in this room has fourteen horizontal slats. But why would they? Unless the number of slats in my window blind or the location of my copy of *The Intentional Stance* is a topic of conversation, then there's no reason to ascribe this belief to me. I could, in principle, be ascribed both beliefs based on the fact that the window blind is in my field of vision and a bookshelf is right beside me. Sensory confrontation is the "*normally sufficient"* (Dennett 1987a, p. 18) condition for ascribing beliefs, and my senses have been confronted with

both of these facts now for some time. But those beliefs would never be ascribed unless they became *relevant* (say, if someone asked me where my copy of *The Intentional Stance* is or how many slats does my window blind currently have). This is a point that Dennett makes clear in "True Believers":

> What we come to know, normally, are only all the *relevant* truths our sensory histories avail us. I do not typically come to know the ratio of spectacle-wearing people to trousered people in a room I inhabit, though if this interested me, it would be readily learnable. It is not just that some facts about my environment are below my thresholds of discrimination or beyond the integration and holding power of my memory...but that many perfectly detectable, graspable, memorable facts are of no interest to me and hence do not come to be believed by me. (Dennett 1987a, p. 18)

We *could* ascribe a belief to Dennett about the ratio of spectacles to trousers in a room (based on Dennett's sensory exposure), but why would we unless that ratio was relevant? If we *do* ascribe that belief to Dennett based on his sensory exposure, then it's because the ratio is relevant for some reason; if we don't, then it's because the ratio isn't relevant. A question about the ratio, for example, would give the ratio relevance, but nothing changes for *Dennett* whether the ratio becomes relevant or not. His visual field contains the same information whether or not the ratio is made relevant. The question is, what is it about Dennett that allows me to ascribe this belief to him? He *must* be able to make visual discriminations, for example, and count, and know what spectacles, trousers, and ratios are in order for him to answer the question. If we can ascribe a belief to Dennett about the ratio of spectacles to trousers in a room, then he must be able to see features,

count, and so on. An ascription of this belief *requires* that Dennett has these capabilities.

Not only must Dennett be able to count, but he must also be able to find features in his environment *interesting* or *important enough* to warrant discrimination. In other words, ascribing a belief about the number of slats, or the ratio of spectacle-wearers to trouser-wearers, requires more than sensory exposure because only *salient* beliefs are ascribed using the intentional stance (as seen in the quote on the previous page). In addition to Dennett's capacity to find environmental features interesting or important, we (as intentional stance adopters) must be able to *discern* that these features are interesting or important to Dennett. What makes a particular belief a salient ascription is, according to Dennett's view, dependent on context and circumstance. If a question about the ratio of spectacle-wearers to trouser-wearers is part of the conversational context (e.g., someone is asked a question about the ratio, thus making the ratio salient, and of interest to Dennett), then the ratio is relevant. So, in addition to sensory exposure, context and salience both play an important role for a particular belief ascription.

Behaviour, contextual salience, and sensory exposure aren't the only capabilities affecting belief ascription. I introduced this particular capacity in the previous paragraph; we can find confirmation of this capacity in Dennett's own words, from the passage in "True Believers" that I quoted above on page 32: the idea that I can notice facts (such as the ratio of spectacle-wearers to trouser-wearers in this room) based on what interests me. Compare two scenarios; one in which Dennett *is* ascribed a belief based on the ratio of spectacle-wearers to trouser-wearers, and one where Dennett isn't ascribed that belief. What must be different in these two scenarios, given that in both cases Dennett has received the same sensory exposure? What is the difference between something

interesting and something uninteresting? I've already mentioned context, but there's something more. That something is related to the capabilities that Dennett has as an intentional system, and can be understood to be something like *attention*. Attention,[11] here, refers only to the capacity for Dennett (and other animals) to pick out or focus on features[12] of his experience. Behaviour, contextual salience, and sensory exposure aren't the only influences for the intentional stance's ascriptions; *attention* makes some sensory exposure salient, in some contexts.

Whether Dennett is ascribed a belief based on the ratio in this example will depend not only on context but also on Dennett's own ability—as an intentional system—to attend to various features in his environment. Some environmental features are ignored, and are thus not subject to attention; those features wouldn't be part of a belief ascription's content. Maybe Dennett *did* compute the ratio of spectacle-wearers to trouser-wearers but kept that information private. Examples such as this, where systems keep information *private* by not publicly expressing a belief based on that information, aren't always interpretable using the intentional stance. Some people are more observant than others, but mostly, different people observe different things about their environment. I will explore this point further in Chapter II, but for now, this point—that a system detects only a small percentage of the features in its environment—is *built into belief ascription* because otherwise there would be no limit to the beliefs ascribable to that system.

In other words, there must be some way for a system to pick out features of its

[11] The specific mechanisms realizing attention are not important for my present purposes, the only important aspect is that intentional systems have this capacity.
[12] These features can range from visual or aural experience, or even features such as pain, thirst, or anxiety.

environment to the exclusion of other features. Systems simpler than humans, such as a chess-playing computer program, have very limited environments (the state of the chess board, along with the moves made by the program's opponent, serve as the chess program's "environment"[13] in this case), so issues involving the selection of relevant features doesn't come into play. Humans are a lot more complicated than chess programs, and we have many more ways to interact with our environment than chess-programs do (or than bees do, for that matter). It doesn't matter *how* this process (of *attention*, or *selection*) is accomplished for our present purposes; it only matters that we *have* this capacity. This particular capacity is another example of how the brain is connected to psychological attitudes.

So far, I've introduced two different examples of how the intentional stance requires its subjects to have certain capabilities, that are required by the psychological attitudes ascribed to that system by the intentional stance. The first example involved looking at the bee's ability to have her behaviour predicted or explained using an attitude even though we have no idea why the bee is behaving the way she does. Our ascription of *wanting to find nectar* is based on the bee's behaviour, but without the right tools, we can't know that this behaviour is, in part, a result of a particular capacity to see UV light. The behaviour caused by the brain is caused by content that we can't possibly ascribe to the bee, even though we can describe the consequences of this content that are made visible in the bee's behaviour.

The second example is that only *salient* attitudes are ascribed to intentional systems. What makes an attitude salient is, according to Dennett, sensory exposure, and

[13] When I use quotation marks, I'm using them as "scare quotes" unless I'm citing the title of a book chapter, a journal article, or a direct quote.

interest. But there is also another influence for a belief ascription, namely the capacity of an intentional system to *pick out* some features of its environment as more interesting than others. One of the constraints on belief ascription involves this *picking out*: an impossible number of beliefs can be ascribed to any intentional system at any time, and the only way to figure out which beliefs will actually predict or explain the system's behaviour is to ascribe only *salient* beliefs. A belief is salient because the content of the belief is interesting or important to the system. Attention is an example of a capacity allowing for the system to discriminate interesting or important environmental features.

These two examples illustrate how the intentional stance depends on its subject having certain capabilities for responding to its environment. Sensory (or perceptual) capacities, as well as the capacity to *pay attention* to one feature of the environment over another, are all required for the intentional stance to work. If a system didn't have any sensory or perceptual capacities, then it wouldn't be able to gather any information about the environment. Humans and bees can, for instance, respond to environmental features using visual discriminations. A chess program also responds to its "environment",[14] although in a way completely different from humans, bees, or other intentional systems. Depending on what's happening in the game, a chess program may choose to pull its queen out early. *What's happening in the game*, at the time the program chooses a move, is the context of the move. In the same context, the program will likely not advance its pawns, or move its rooks, because this context demands that the queen should be pulled out early. Pawns and rooks are ignored in favour of getting the queen out early in this situation.

[14] As I claimed on page 35, a chess program's environment, in this case, is the state of the board in the game the program is currently playing.

Sensory and perceptual capabilities are both examples of design features that a subject must have for an intentional stance prediction or explanation. Without these capabilities, the subject wouldn't even be able to behave. Dennett accepts that an intentional system must have capabilities that are a consequence of its design, but he doesn't explore the further consequences of these capabilities for the intentional system. Some of the capabilities of an intentional system, as I will argue in this project,[15] can help explain the basis on which systems can form psychological attitudes.

For now, however, I want to continue focusing on the intentional stance's reliance on its subject's capabilities. Assuming that a system is evolutionarily designed, and will function as designed, will bestow certain capabilities on that system. If a subject can be ascribed a belief based on detecting UV light, then the system must have the capacity to detect UV light. That being said, not only does the intentional stance assume that its subject will function as it was designed, but the stance also assumes that its subject is rational. Assuming that a system is rational obviously requires that the system has the capacity to behave rationally, and in the next section I will investigate the consequences of the rationality assumption adopted by the intentional stance toward its subjects.

1.3. Capabilities related to the rationality assumption

Whereas the design assumption bestows particular capabilities on designed systems, the rationality assumption assumes that the design system will respond

[15] Chapters I, II, and III all deal with the capabilities that an intentional system (in particular, humans) has allowing it to be able to form psychological attitudes. The critical capability for humans, is the ability to have mental representations about its experience; these representations are specifically the subject of Chapter II.

appropriately to its environment. Here, *appropriately* means *rationally*.[16] If you believe that the stove is hot, and you desire to avoid first- or second-degree burns, then not putting your hand on the burner is an appropriate response. An evolutionarily designed system is designed to avoid the harms it can recognize. Likewise, the intentional stance's assumption of design includes an assumption that the designed system is also rational and will avoid harms it recognizes. I'm sure that no one will disagree that avoiding harm is an appropriate thing to do, given that harm can be lethal. *Avoiding harm* is an obvious example of the control rationality asserts over our behaviour.

What does the rationality assumption mean for an intentional system and its capabilities? In this section, I will investigate an intentional system's capabilities related to the system's rationality. The brain of an intentional system is connected to the attitudes that system is ascribed because the attitudes are ascribed based on behaviour—to the brain's output—and that behaviour is rational, i.e., appropriate for its situation. But there are consequences of the rationality assumption for intentional systems. I will use Dennett's (1987b) example involving Sherlock, Jacques, Tom, and Boris, illustrating that two different systems (at least two) may be independently ascribed the same belief based on completely different reasons. Because all of the actors in Dennett's example is rational, their behaviour can be predicted by ascribing all of them the same belief.

Each actor is rational—and thus their behaviours are rational—so the capabilities that each actor has are suited to produce rational beliefs as a response to their environment. Their *beliefs* are rational because they rationally cohere with the rest of their beliefs (and desires). A system's rational responsivity allows us to ascribe the same

[16] I discussed the rationality assumption in section 2 of the Introduction to this project. I will not repeat what I wrote there.

belief to two different systems based on context and environment. The intentional stance might attribute the same belief to two different subjects based on very different experiences. Two different systems, with entirely different behaviours and experiences, sometimes deserve the same belief ascription because of the way that particular ascription explains or predicts their behaviour. Dennett's example of Sherlock, Tom, Boris, and Jacques from "Three Kinds of Intentional Psychology" illustrates this point:

> Jacques shoots his uncle dead in Trafalgar Square and is apprehended on the spot by Sherlock; Tom reads about it in the *Guardian* and Boris learns of it in *Pravda*. Now Jacques, Sherlock, Tom, and Boris have had remarkably different experiences—to say nothing of their earlier biographies and future prospects—but there is one thing they share: they all believe that a Frenchman has committed murder in Trafalgar Square. (Dennett 1987b, p. 54)

All of the characters in this illustration (except for the uncle, of course) can be ascribed the belief that *a Frenchman has committed murder in Trafalgar Square* based on different reasons. The disparate experiences of Jacques, Sherlock, Tom, and Boris can all lead to the same belief ascription because each actor can have their behaviour explained using a belief that would be rational for each to have, given their experiences.

The distinct experiences of Sherlock, Jacques, Boris, and Tom in Dennett's example illustrate how a belief ascription is dependent on information the subject is presumed to possess, given the subject's experiences. Any of those people might express the belief that *a Frenchman committed murder in Trafalgar Square* (F) if asked. If Boris and Tom knew Jacques' uncle, then they may even go to the uncle's funeral; this behaviour could be predicted or explained by ascribing F to them. They may each have

different experiences surrounding their expression or ascription of that belief, but the same belief can be ascribed based on both Tom and Boris' behaviour.

Even though all four people in the example do not share the same experience, they must share something in common, otherwise, on what basis could they all be ascribed the same belief? One thing that they share in common is how their experiences affect their behaviour. They all possess certain information about the world (there was a murder) gathered from different sources (reading a paper or apprehending the murderer on the spot). They *don't* need to share a discrete brain state corresponding to that belief. Instead, there are patterns in their differing experiences used as a basis for an ascription of the same attitude to each person.

Where are the patterns in this case? Tom and Boris each read a different newspaper with a similar story on the murder in Trafalgar Square. Their exposure to the story was by reading a newspaper, and this exposure is the basis for the belief ascription. *Exposure*, here, is *visual exposure* and refers to Tom and Boris's *reading* of the newspaper. If Tom and Boris *heard* from someone about this particular murder, we'd also base a belief ascription about the murder in Trafalgar Square on their aural experience; *hearing* can also expose one to information. In these cases, whether Tom and Boris *read* or *heard* about the murder doesn't matter, what matters is their exposure to that information. Knowledge about exposure to information will help us figure out where patterns can be detected; we would adjust our intentional strategy toward an individual who was either blind or deaf because the blind or deaf person is missing some of the faculties often warranting a belief ascription.

To summarize the previous paragraph: experience plays an important basis for

beliefs in IST, "exposure to *x*, that is, sensory confrontation with *x* over some suitable period of time, is the normally *sufficient* condition for knowing (or having true beliefs) about *x*" (Dennett 1987a, p. 18). In IST, ascribing a belief about *A* to *x* because *x* had some sensory confrontation with A is a common reason why *x* is ascribed a belief about A. I have argued in the preceding sections that patterns in behaviour are—ultimately—the result of a system's evolutionarily designed capabilities, such as sensory capabilities and an ability to rationally respond to experience. In other words, the behaviour exploited by the intentional stance is a response to experience.

Thinking of behaviour *as a response to experience* is important because behaviour is an output of the brain. Since the brain belongs to an intentional system, the system will be behaving, for the most part, rationally. That is, the system will respond appropriately to her situation (such as finding food or hiding from predators) given what she believes and desires. If we asked Boris or Tom about whether or not something happened in Trafalgar Square, we could predict their answer based on the belief they've been ascribed. Tom and Boris are both rational systems, and if we knew that they each read a newspaper the morning after the murder, then it would be very likely that Boris and/or Tom would say that he remembers reading something in the newspaper about a murder last night in Trafalgar Square. Very generally, Tom and Boris are rational systems, and the behaviour output by their brains will also be rational. The intentional stance depends on this rationality when making ascriptions based on behaviour or experience.

People can have their behaviour predicted or explained using the intentional stance because their brains produce the patterns in their behaviour. A person acts in a

predictable or explicable way, displaying a detectable behavioural pattern because that person is rational and that person's brain will produce rational behaviours. That brain is evolutionarily designed, and is assumed to function as designed, so the capabilities it has as a result of this evolutionary design will be related to its behaviour. Patterns in behaviour connect an ascribed attitude to the brain because those patterns provide insight into what's going on inside that brain.

We can get a lot of information about what's going on inside the brain of an intentional system by basing our ascriptions on their behaviour. Brains have many different capabilities, and those capabilities are on display when we ascribe psychological attitudes based on the behaviour that's part of the display. The brain is *displaying* its contents using behaviour because of the direct causal link between brain and behaviour. By ascribing attitudes based on patterns in behaviour, the intentional stance also, indirectly, ascribes attitudes based on something going on inside the brain. The next section will investigate this *display* and its relationship to the intentional stance.

1.4. Displaying brain content behaviourally

The assumptions of rationality and design assume that intentional systems are designed to handle inputs and outputs appropriately. *Inputs* and *outputs* can be in many different forms; for example, the brain takes sensory information, such as information gathered by vision or hearing, and uses this information to produce outputs such as behaviour in response to those inputs. In other words, what the brain can do is on *display* in behaviour. The intentional stance ascribes various attitudes based on different parts of this display. Using this stance, we might say that *x* stops at a traffic light because *x*

believes that the light is red, and *x* doesn't want to risk a collision. *X*'s behaviour displays both of those attitudes.

In this section, I will examine how this display is one of the brain's responses to the world. This display is a result of a system's capabilities involved with responding to the world, and is, among other things, also a display of the system's capabilities. The brain displays its capabilities, and its capabilities also influence its display. For instance, *x* would stop at a red traffic light because *x* sees that the light is red and knows the rules of the road. Another person might ascribe a belief that *the light is red* to *x* based on *x* stopping at the red light. *X's* evolutionary design allows him to respond to visual stimuli, and *x* is also rational, so he should have rational responses to visual stimuli.

An intentional system responds to its environment rationally, and with this rational responsivity comes certain capabilities. The content of an ascribed attitude, for example, requires that the bearer of that ascription has certain capabilities associated with the content. The examples from section 1.2 and 1.3 illustrate some of these capabilities. For instance, Tom and Boris must be able to *see* and *read* to have a belief about the murder, so when an ascription is made to Tom or Boris based on their experience of reading a newspaper, that ascription is made with the assumption that they can see and read.

In general, how subjects *respond* to their experiences provides some insight into the subjects themselves. But not all possible responses actually make it into the display. In IST, a system's intentional features are those features attempting to explain the display in the system's behaviour. Behaviour is displayed, and by explaining behaviour using psychological attitudes, the intentional stance is attempting to explain what the brain is

presently displaying. Psychological attitudes are attributed based on the visible patterns in behaviour constituting this display, but IST's analysis cannot explain how a subject forms her own attitudes without that subject forming attitudes based on an observation of behaviour.[17] Patterns in Tom and Boris's behaviour can be the basis of an ascription of F, but that pattern would exist even if no one was around to make an ascription. Just because a pattern is indiscernible from the intentional stance doesn't mean that the pattern doesn't exist.

To summarize the chapter thus far: a particular belief or desire, such as a bee's desire to find nectar or Tom's belief that a Frenchman committed murder in Trafalgar Square, can predict a subject's behaviour because that behaviour constitutes a display of the subject's brain. For instance, if a subject can be successfully predicted using a belief involving *seeing* an eagle, or *reading* a newspaper, then that subject must have certain capacities involved with *seeing* and *reading*. In the case of seeing or reading, vision or literacy are (usually) at the very least required for those attitudes. Both the monkey's vocalization and Tom or Boris's answer to a question about Trafalgar Square are displays of what's happening inside each brain.

What a subject can display is evidence of the subject's capabilities. The intentional stance's requirement that a subject displays her mental contents means that certain ascriptions require the subject's brain to be designed in a certain way. In the next section, I will examine the implications for an intentional system's design in light of the

[17] I'll reiterate this point; according to IST, if the intentional stance attributes attitudes based on mostly behaviour, then the attitudes we e*xpress* must also be self-attributions based on an observation of our behaviour. In other words, the attitudes we form (whether the attitude is privately articulated or expressed, for example) are the result of us interpreting and ascribing those attitudes by observing our own behaviour in this view.

considerations from sections 1.2 to 1.4 involving the intentional system's display of its capabilities.

1.5. An intentional system's design

This section will investigate the consequences of an intentional system's capabilities for its design. I intend to show that if we fully consider the consequences of an intentional system's capabilities, then we'll begin to see how psychological attitudes can be formed by that system without observing its own behaviour as the attitude's basis.

Officially, design, and how that design is implemented, is not the subject of IST except for an intentional system's propensity to have beliefs and desires. Dennett calls IST and the intentional stance "black box" theories (for example, see Dennett 1987a or 1987b). "Black box" theories are theories that ascribe "beliefs (and other mental features) by behaviour, cultural, social, historical, *external* criteria" (Dennett 1987a, p. 14). External criteria are used to judge whether or not a person has a particular attitude, and these externally discernable features are *all* that the intentional stance can access to base its psychological attitude ascriptions on.

Additionally, patterns in behaviour are ascribed attitudes based on cultural, social, historical, or experiential information. For instance, Tom is an attentive reader, and he always reads local crime stories in the paper, so ascribing the belief that *a Frenchman committed murder in Trafalgar Square yesterday* would have not only a rational basis but also a basis in Tom's biography. There's nothing directly detectable inside Tom's head helping this attribution from the intentional stance's point of view. The only detectable things are what Tom displays in his behaviour. In other words, Tom's brain is a black

box.

Black box theories such as IST and the intentional stance are supposed to be neutral regarding how the brain realizes or implements a psychological attitude (Dennett 1987a, p. 24). The subject matter of both IST and the intentional stance isn't supposed to include anything about the brain. *Internal* features (features involving the brain's processes, in contrast with external features) are by definition ignored by the intentional stance as a basis for psychological attitudes. Questions about internal features are secondary considerations, because "before we ask ourselves how mechanisms are designed, we must get clear about what the mechanisms are supposed to (be able to) do" (Dennett 1987b, p. 74). IST deals with ascriptions; only external features are used to base an ascription. One of the reasons that IST ignores internal features is its assumption of design. This assumption puts any questions about design implementation aside and instead focuses on what the system is designed to do.

But there are only so many ways that a system could possibly implement what it was designed to do. There may be a very large number of ways, but there is still a limit on how an intentional system may be implemented. A system that can be ascribed a belief based on its visual experience, for example, must be designed for responding rationally to visual stimuli. For example, ascribing a belief to a monkey based on its vocalization when it sees an eagle assumes that the monkey can both see and vocalize.[18] If the monkey can see the eagle, then there is some content in the monkey's head representing the eagle allowing for the monkey to respond.

This claim that the monkey's brain represents the eagle raises some interesting

[18] See section 4 (pp. 13-14) of the Introduction for the example of monkey's vocalization abilities.

questions. For instance, *how* does the monkey's brain represent the eagle? Is there a *type* of brain state in the monkey's head that's identifiable as this particular belief? According to IST, the answer is a resounding *no*. For Dennett, types of psychological attitudes are *not* reducible to types of brain states; psychological attitudes are ascribed to subjects based on patterns in that subject's behaviour. This rejection of type identity theory is important because of the role that the brain's representational powers will play in the account of psychological attitudes that I will provide in Chapter II. Since I'm working within Dennett's framework, type identity theory is rejected in the account of psychological attitudes that I will provide in the chapters to come.

IST adopts a different basis for the monkey's belief *that there's an eagle in the sky* from a basis that involves the monkey's brain. The eagle may not be represented in the monkey's head as a discrete type of brain state, but there is one way the eagle *is* represented: patterns in the monkey's behaviour. The monkey's vocalizations and gestures are a representation of the eagle for the purpose of alerting the rest of his troop. The monkey is acting in such a way as to alert his troop of an eagle approaching from the east; when the rest of the troop sees and hears the first monkey's vocalizations and gestures, they know that there's an eagle approaching from the east. The troop can associate the gesture/vocalization with an eagle approaching from the east (for example) because the gesture/vocalization represents an eagle approaching from the east.

In other words, one way the eagle can be represented is in the monkey's behaviour. The monkey's behaviour can be interpreted as representing his experience of the eagle by a researcher using the intentional stance: the belief ascription of *there is an eagle in the sky* is another way to represent the information the monkey must possess,

given the belief ascription. In this case, the monkey's behaviour is represented by the researcher using a belief with particular content (*that there is an eagle in the sky*).

The brain *doesn't* represent attitudes as discrete types of states; a researcher *can* represent a subject's behaviour using an attitude based on that behaviour. A belief may not exist as a type of brain state, but it does exist as an ascription based on behaviour. This is the position of the intentional stance and IST.

That being said, behaviour—as I have argued in this chapter—is the result of a brain reacting to its experience so there must be *some* connection between the brain and its experience. That connection *doesn't* have to be realized as a *single, discrete brain state*, in the same way that the approaching eagle isn't represented by the monkey with a *single motion*. Just as a whole series of motions are involved with the monkey raising his arm to gesture, and flexing his vocal chords in such a way to produce a vocalization, the eagle can be represented in the brain with multiple representations. How the brain accomplishes this will be discussed in Chapters II and III; for now, the point I want to make is that the brain must represent its experience in some way. I am claiming (uncontroversially) that the brain controls behaviour, and this behaviour is a response to the world, so the brain must be controlling the responses in some way. If the brain responds to its experience using its evolutionarily designed capabilities—sight and speech, for example—then the brain must possess information derived from the world using those capabilities.

A system's capabilities allow for the system to react in certain ways to certain things; a simple example is a visual experience causing a person, a bee, or a monkey to behave rationally according to that experience. A bee can find nectar using UV light, a

person can read a newspaper and answer questions about stories in the newspaper, and a monkey can inform its troop that an eagle is in the sky. A brain must be able to represent its environment somehow. Otherwise, it wouldn't be able to produce behaviour reacting to that environment. We can find a healthy medium between representing psychological attitudes as discrete types of brain states and representing those attitudes using an ascription based on behaviour. This medium can be in the form of mental representations: there must be mental representations of visual experience (for example) because not only can the brain react to events using visual experience, the brain can react to *past* visual experiences.

Those mental representations don't have to be discrete types of brain states. There doesn't have to be a type of representation for belief, or desire, or fear, but there does have to be a representation of what's believed, desired, or feared, in order for the brain to control the behaviour that ascriptions of belief, desire, or fear are based on. The brain is designed to represent the world and act according to those representations. This design includes capabilities, capabilities that allow for not only behaviour in general, but specific behaviours as a response to experience. In other words, looking at an intentional system's designed capabilities can provide insight into how a brain represents the world and itself.

In short, examining an intentional system's capabilities will provide a place to look for a source of the psychological attitudes that we ourselves form. In this section, I attempted to claim the intentional stance requires a system to be designed in a certain way for that system's behaviour to be predictable. If we're clear about what the mechanisms (involved with an intentional system's brain) are supposed to do, to use Dennett's phrasing, then an intentional system must be able to do specific things in order

to be ascribed a psychological attitude. Here's what I mean. If a system is capable of having its behaviour predicted using an attitude, then the design of that system must be able to realize the capabilities responsible for its behaviour—behaviour the intentional stance bases its attitude ascriptions on—using representations of some form. For instance, a system whose behaviour can be predicted by ascribing the belief that *a Frenchman committed murder in Trafalgar Square* must be able to gather information related to that belief. Seeing, reading, memory, etc., are all examples of capacities that are required for a system to have its behaviour predicted or explained using that belief. That system must be representing the content of the belief somehow in order for him (or her) to behave according to that belief.

1.6. Conclusion

The intentional stance relies on an intentional system to display its mental contents for predictions and explanations of that system's behaviour. These mental contents are the result of the system's capabilities and manifest as patterns in behaviour. A bee, for instance, can detect and respond to the UV spectrum by behaving in ways that can be the basis for an ascription such as *hunting for nectar*.

In sections 1.2 and 1.3, I argued that the intentional stance requires that an intentional system has particular capabilities, and these capabilities are the basis for the patterns in behaviour used by the intentional stance. Not surprisingly, behaving because of something you saw or heard, for example, is the result of being able to see and hear. This point is obvious, but it underlies many of our more complicated behaviours. Both Tom and Boris can be ascribed the same belief that F, despite their different experiences

and different behaviours, because of what we know about Tom and Boris's capabilities as intentional systems. The intentional stance uses these capabilities to make its psychological ascriptions, because of the way those capabilities influence behaviour.

Section 1.4 examined how an intentional system displays these capabilities. The intentional stance relies on this display and the various capabilities causing behaviour that constitute this display. Much of what's expressed in this display is a result of these capabilities—and thus relied on by the intentional stance—but these expressions aren't the result of the system observing its own behaviour. These behaviours are the result of something that's connected to what the brain can do. For instance, the stockbroker's picking up of the telephone because he wants to buy some stocks, a bee using UV light to find flowers; both of these examples illustrate the fact that an intentional system's behaviours are based on the capabilities that system has for responding to its experience.

The final section, 1.5, argued that an intentional system must be designed in such a way as to produce these behaviours. These behaviours are a response to the world, and there must be some kind of internal feature representing the world (derived from the system's capabilities) that allows for this behaviour to exist in the first place. These internal features—features that *aren't* observable behaviour but *are* a result of the system's capabilities—must be the basis for the psychological attitudes that we ourselves form. The next chapter will examine the internal features that I claim serve as the basis for many of our psychological attitudes.

Chapter II.

The Representational Content of Psychological Attitudes

2.1. Introduction

In sections 1.2 and 1.3 of the first chapter, I argued that intentional systems have certain capabilities (such as the ability to see or be literate) resulting from their evolutionary design. In the Trafalgar Square example from the previous chapter, Tom or Boris must be able to *read* for the belief *a Frenchman committed murder in Trafalgar Square* to make any sense. These capabilities may be obvious, but they're critical for the intentional stance. Behaviour is a *display* of the content gathered by some of these capacities, and the intentional stance depends on this display.

The importance of recognizing that the intentional stance depends on its subject displaying its brain contents is that some of this content doesn't have to be expressed. For example, usually keeping a secret requires that you don't let the secret affect your behaviour. Unless you're notoriously bad at keeping secrets, that secret won't affect your behaviour. Non-public attitudes, such as a secret, are familiar examples of attitudes that don't affect behaviour. These beliefs are *privately articulated*[19] but not publicly expressed. Not only do we have the capacity to express our beliefs, but we also have the capacity to articulate beliefs to ourselves privately.

Not only are these privately articulated attitudes (such as a secret) unascribable by the intentional stance, but we couldn't form those attitudes by making inferences about our behaviour. Basing *x*'s articulation of belief *p* on inferences about behaviour means

[19] I introduced my notion of *private articulation* in section 4 (pp. 14-15) of the Introduction to this project. An attitude that's kept private and not publicly expressed is a private articulation.

that *p* wouldn't be private. There would be no secrets if behaviour was the basis for all psychological attitudes. An observer could ascribe an attitude based on the same behaviour that *x* uses for his inference. What is the basis for *x*'s private articulations?

A theory about psychological attitudes that doesn't give behaviour the primary focus seems to be at odds with IST. I think the opposite is true: the abstract, pattern analysis of intentionality provided by IST is not only that theory's most important contribution to the understanding of intentionality, but this contribution can also be extended to include sources of information other than behaviour. This chapter will argue that *representational content*[20] is the basis of the attitudes we form ourselves.

So far, my supplement goes like this. Intentional systems have capabilities granted to them by the assumptions of design and rationality. Here are two examples of these capabilities. First, intentional systems have capacities for collecting information about the world. Humans (and other animals) can gather information from a wide variety of sources, such as colour vision or the regulation of water levels in the body. Second, some intentional systems can display the information they possess in behaviour, such as a bear's intimidating roar or a human ordering lunch at a restaurant. We might not know what a bear precisely expresses, but we can get a pretty good idea by watching his behaviour (go away or I might eat you!).

Obviously, these two examples are not exhaustive. There is more to my supplement of IST than simply placing emphasis on our evolutionary design and rationality. There must be other capabilities giving us the capacity to form our attitudes: we do it all the time! In this chapter, I'll argue that our brain's ability to represent

[20] See section 5 (pp. 16-18) of this project for my introduction to representational content.

information serves as the foundation for the psychological attitudes we form. More specifically, this chapter is about the representational basis of our psychological attitudes and our relationship to that basis.

The first section, 2.2, will argue that there are multiple sources of information that our attitudes are based on. Psychological attitudes can be based on events external to ourselves, using our sensory faculties such as vision, or an attitude can be based on internal features. Memories or internal sensations associated with attitudes like fear and anxiety are both examples of information serving as a basis for these attitudes. Notice that neither of these sources exclusively involves observing our behaviour.

Section 2.3 argues that representational content is ultimately the basis of many of our attitudes. We can see features in our environment, or remember attitudes that we've formed in the past, but in any case, this information is somehow represented in our brains. Not only can we form and display our mental representations, but we can also form and privately articulate those representations using an attitude.

The difference between expressions and ascriptions is the subject of section 2.4. Representational content can serve as a direct basis for a psychological attitude expression, but only when that content is displayed in behaviour will representational content influence an attitude ascription made by others.

In the final section, 2.5, I explore Dennett's notion of a theorist's fiction and how it affects the psychological attitudes we form. In this view, our psychological attitudes consist of judgements about the stimuli causing our representational content. As a result, our attitudes can be considered a *theory* regarding what we're experiencing. Our attitudes based on this content constitute something like a fictional account of our experience in

Dennett's view. Truths can be discerned from this fictional account similar to the way that truths can be discerned from works of literary fiction, according to Dennett, regardless of the judgments our attitudes are based on accurately reflect the world.

2.2. The non-behavioural sources of content for attitudes

In this section, I'll outline the sources of the information that forms the basis of our psychological attitudes. These sources range from our capacity to gather information about our environment to our capacity to respond to physiological changes. We can even form attitudes from other attitudes. I'll argue that understanding the sources of our information will provide insight into how we form these attitudes.

Our capacity to gather information about our environment plays a major role not only in our attitude formation but also in the ascriptions based on our behaviour using the intentional stance. For instance, I'd stop at a traffic light when the light is red. I can *see* that the light is red, and that visual experience would affect my behaviour as well as allow me to privately articulate or verbally express a belief that the light is red. The reason that another person would explain my behaviour (using the intentional stance) at the traffic light is similar to the reason I'd use to explain my behaviour in an identical situation. The other person would assume that I stopped because I can see the traffic light. *The traffic light is red* is a belief that I ought to have, and is the kind of belief that the intentional stance would ascribe to me based on my behaviour.

There's a lot of information available to base a belief ascription on. Recall Dennett's most basic, minimal definition of belief, which will be important for the rest of this project:

The information causing an agent to act. (1998b, p. 324)

I see the red traffic light, ergo I stop. Sensory experience (seeing the red traffic light) is one of the reasons I stop. However, sensory experience alone cannot account for all of our attitude expressions; for instance, my apartment is on the second floor of an apartment complex, I walk upstairs to the second floor not because of my sensory experience, but because I remember that I need to go to the second floor in order to go home. My belief that my apartment is on the second floor causes my behaviour in this case, but that belief isn't the result of immediate sensory experience. My behaviour, in this instance, is caused by my memories of my apartment's location.

What I'm currently experiencing and what I have experienced are both important sources of information for our attitudes because we frequently form psychological attitudes about both our present and past experience. However, sensory history is not the only source of our attitudes. This point should be evident: the intentional stance ascribes attitudes based on more than sensory history. A system's other beliefs, for example, or that system's memories and biography, are all equally important sources of information that may warrant different attitude ascriptions. Similarly, we form psychological attitudes based on our memories, how we're feeling, or even other attitudes. In other words, senses aren't the only basis for our attitudes. Dennett's minimal definition of belief doesn't qualify the information motivating us to act; a psychological attitude can be based on information derived from a wide variety of different sources.

Humans (and other animals) are endowed with certain sensory capacities for collecting information about their experience. An organism with visual faculties is capable of receiving visual sensations, for example, which are in turn perceived and

given meaning. Our sense faculties are a major source of information for our attitudes.

Sensory experience is a requirement for many of our attitudes, but not all of this information can be the basis of an attitude. In Chapter I, I referenced Dennett's claim that the intentional stance ascribes beliefs based on only salient details. The information used to perceive and calculate the ratio of spectacle-wearers to trouser-wearers might be part of our visual field, but unless that information was of interest, we wouldn't form a belief involving the ratio. Not everything sensed is the content of an attitude, but in order to form attitudes about features of our experience, those features must be subject to our sensory faculties.

There is an important, and very relevant, distinction within the category of *sensation*. The distinction is between *interoception* and *exteroception*. Exteroception is the reception of external stimuli, and vision and auditory faculties are excellent examples of exteroception. Interoception[21] is the reception of information from the body; temperature, hunger and blood pressure are examples of interoception (Pinel 2011, p. 174). Our brains have the capacity to sense information from both internal and external sources. This distinction is important because interoceptive sources of information are invisible to the intentional stance unless that information is displayed. Interoceptive sources can be the basis of private articulations; if I've a feeling of dread or anxiety that doesn't affect my behaviour, the intentional stance won't be able to ascribe either of these attitudes to me. I can have this attitude, even though the attitude won't be predicting or

[21] It's important to note that there is a difference between *interoception* and *introspection*. Interoception is just the brain's ability to respond to internal events. *Sweating* when it's too hot outside is an example of the brain regulating body temperature, which is an interoceptive mechanism. *Introspection*, on the other hand, is "looking into one's own mind" (Gregory 2004, p. 485) and is an active process involving things like awareness or consciousness of what's going on in our mental life. Interoception is a brain mechanism and considered a form of sensation; introspection is not a form of sensation or perception. Introspection will be discussed later in this project.

explaining my behaviour from an intentional stance.

So, very generally, there are at least two separate systems of gathering information: internal and external systems. That being said, not all of this information can be the content of a belief; beliefs about sensory information that *could* be formed often aren't because those beliefs are trivial. Why would I verbally express the ratio of spectacle-wearers to trouser-wearers?

When the intentional stance considers its subject's sensory history, the stance is making inferences involving the subject's external stimuli. Internal stimuli—via interoception—is available to the intentional stance when those stimuli are displayed in behaviour. Stimuli that aren't displayed remain unavailable to the intentional stance unless they end up influencing behaviour.

Many of our beliefs, whether or not those beliefs are formed, could be ascribed to us by the intentional stance based on sensory and perceptual capacities. For example, in the Martian illustration from Chapter I, section 1.2, the Martians can see the many different bodily motions but cannot explain the patterns in those motions using a psychological attitude because they don't have the intentional stance. The stockbroker could be attributed any number of beliefs based on his various minute bodily motions constituting his behaviour even though none of these beliefs are of the kind that people would usually ever form. Similarly, Dennett would never form a belief about the ratio of spectacles to trousers unless that ratio became salient (say, if Dennett was asked a question about the ratio of spectacles to trousers in the room). The Martians are in a similar position as Dennett toward the ratio of spectacles to trousers: the patterns are *there*, whether or not those patterns are discernible. Psychological attitudes are of no

interest to the Martians (because they don't know those attitudes even exist) just as the ratio is not interesting to Dennett.

Here's another example of information that can be the source of an attitude that isn't derived from sensory experience. Other attitudes can cause our behaviour, whether those attitudes involve past sensory experience or not. If I believe political candidate D is not good for the country (1), I won't vote for him (or her). I may remember past sensory experiences, such as hearing or seeing D speak at a rally, which might influence forming the belief in (1). However, I may have also interpreted D′s political dialogue, weighed that dialogue against other political and economic issues, and come to the opinion that (1). I don't base a self-ascription of (1) on an inference about my behaviour; the inference is based on my other attitudes. Clearly, we can form attitudes that aren't just based on sensory experience.

In fact, many of the attitudes we form aren't based on sensory experience at all. For instance, a system which expresses anxiety about an event isn't basing that attitude on its own behaviour. Instead, that particular attitude is based on something that may be described as *a feeling of anxiety*. That feeling might be caused by an experience such as a scary dog, along with various physiological reactions (such as an increased heartbeat). No matter what the attitude is based on, an attitude of anxiety obviously has *some* basis. That something is not behaviour: seeing a scary dog and becoming scared is not the result of observing your behaviour. It's true that one might notice various current physiological reactions—such as an increased heart rate or light-headedness—and ascribe herself an attitude of anxiety or fear based on those physiological reactions, but most often we simply *get scared* and behave rationally (such as running away, or becoming defensive)

without making any inference, interpretation, or observation of our behaviour. The intentional stance can ascribe a feeling of fear or anxiety to that person based on the person's behaviour, but it's possible that individual will keep her reactions private.

Sensory experience *can* contribute to the attitude's formation; you *see* a scary dog and become fearful, for example. Additionally, an attitude can be formed without sensory experience as seen in the case of (1) from the previous page. If I disagree with state-sanctioned bigotry or racial profiling, then an attitude based on these disagreements isn't based on sensory experience. An attitude of disgust toward a political candidate may find it's source in other attitudes that I've had, rather than anything I see or hear in particular about the political candidate.

The *internal features* provide an important basis for many of our attitudes. IST doesn't develop an account of attitudes without basing the attitudes on behaviour or inferences about sensory experience to any worthwhile degree, and whatever account that can be scrounged from Dennett's writings[22] doesn't provide any serious account of internal features serving as the basis of our attitudes. Our brain may observe[23] its own inner workings, forming the basis for an attitude based on those inner workings (such as of thirst or pain), or even an attitude based on other attitudes. But this mechanism of "observation" is nothing like the intentional stance; there's no assumption of rationality or design involved in this mechanism, for example.

Our psychological attitudes can be based on (this list is not exhaustive) sensory

[22] The passages I quoted from *Consciousness Explained* and "Multiple Drafts Model" (Akins & Dennett 2008) in section 4 (p. 11) of the Introduction serve as a minimal account of psychological attitudes only after heavy interpretation and lots of eye-squinting.

[23] I don't mean that brains observe their inner workings the same way that we observe things visually. I'm referring to cognitive mechanisms that respond to other cognitive mechanisms, or physiological changes; a change in heartrate based on something in the environment, or becoming tired when blood sugar is low are both examples of the kinds of processes that I'm referring to with the term *observe*.

experiences, our other attitudes, or information about physiological changes; in all of these cases, internal features play an important role in the formation of our attitudes. Observing our behaviour as the intentional stance demands isn't required. Instead, the external and internal features can be the origin of the information our attitudes are based on. The next section will argue that this information is represented in the brain, and this representational content is the information serving as the basis for many of our psychological attitudes.

2.3. Representations and core elements

Sensory information is represented in the system's brain but makes up only a fraction of the total number of things that a brain actually represents. Other attitudes and memories are both examples of things that can be expressed or privately articulated (for example) using an attitude, and therefore must also be represented (in some way) within the system's brain. At any one time, mental representations can include a range of information, from the inner workings of the system's body to information about yesterday's events. Most of this information is entirely and forever unconscious (you will never feel your liver producing bile, even though there are representations in the brain helping to regulate this production), and most information will never display itself in behaviour.

In this section, I'll argue that representational content (whether that content consists of sensory information, information about internal features, memories, or other attitudes) forms the basis of the attitudes we express or privately articulate. More specifically, in the following section, I'll outline how those representations fit into my

supplement to IST. My version will amount to a major addition to IST because the intentional stance does not consider internal features for its ascriptions even though many of the attitudes that we form are based on internal features.

Behaviour is the most common basis of attitude ascription for the intentional stance. But we can form our own attitudes that aren't based on interpretations of behaviour. External events may influence my expression of disgust for political candidate D (such as D's bigoted remarks), but my disgust isn't based on *my* behaviour. If anything, this attitude expression relies on *both* external *and* internal features—external features such as D's bigoted words, and internal features such as my propensity to be disgusted by D's words. If I find D's words to be disgusting, whatever my criteria for being disgusted by something are (such as bigoted dialogue) will influence the forming of my attitude. *My own criteria for disgust* are examples of the internal features that I'm referring to.

These internal features—such as my own criteria for being disgusted, other attitudes, or memories—serve as a basis for many of an intentional system's attitudes. The features are represented in an intentional system's brain and compose the representational basis of our psychological attitudes. IST has very little to say about these internal features because the intentional stance uses only external features for the basis of its ascriptions. Mental representations—the internal features—are not available to other people, unless they're indirectly displayed in behaviour. I'll return to this point later in this section, but first I'll investigate what Dennett says about these representations.

Dennett discusses how representations fit into IST in "Three Kinds of Intentional Psychology" (1987b, p. 57) using the name *core elements*. His discussion of the core elements in this paper amounts to very little, and he does not develop the idea any further.

Basically, these elements are the "concrete, salient, separately stored representational tokens" causing many of our attitudes (1987b, p. 57). Mental representations result from a system's capabilities as an evolutionarily designed system, and those representations are both *concrete* and *stored separately*, according to Dennett. That being said, these representations must be something other than just psychological attitudes because attitudes aren't supposed to reduce to separate types of brain states identifiable as discrete attitudes in IST. So what are these core elements, if they're not individualized types of brain states representing types of attitudes?

Core elements are the particular representations forming the basis of our psychological attitudes. The core elements of my belief that *there is a table in front of me*—whether that belief is expressed *by* me or ascribed *to* me—are the representations relevant to the belief. Various sensory information informs a belief about a table. The sensory information is represented, and the portion of information to which our brain perceives and gives meaning (there's a table in front of me) serves as the core elements of that particular belief.

A core element isn't a type of brain state corresponding to a psychological attitude. Instead, *the core elements* (plural) consist of the representations informing the basis of our psychological attitudes. Representations that *don't* serve as a basis for a particular attitude aren't part of that attitude's core elements. These irrelevant representations are analogous to the "noise" accompanying a pattern that I referred to in Chapter I, section 1.2, (p.26n9). Core elements are the basis of my beliefs; I form my belief that *there is a table in front of me* for reasons such as *I can see the table* or *I need an example*. Both the original belief (*there is a table in front of me)* and the reasons I

offer to justify my belief (*I can see the table,* and *I need an example*) have their own core elements.

Dennett says next to nothing about the core elements beyond the very small passage in "Three Kinds of Intentional Psychology" quoted on the previous page, so there isn't much of an account regarding what these core elements are. This is where the gap in this theory lies. If Dennett developed this idea of core elements further, then perhaps he might be able to incorporate internal features into IST's account of psychological attitudes. Since this lack of development presents a gap in his theory, I'll attempt to develop the idea of core elements and fill this gap. The example Dennett uses to illustrate how the core elements are causally related to behaviour involves a person who blushes when she believes another person knows her secret (Dennett 1987b, p. 57). The blushing person doesn't have to have an *explicit* internal representation of her belief that John knows her secret, that particular belief exists only *virtually* (1987b, p. 57). *Virtual* is the opposite of *explicit* in this context: data is *explicit* when the data exists separately and discretely; when data is exists within a *network* rather than individually, that data exists virtually.

Representations in the blushing person's head, in addition to events external to her (John knowing her secret), cause her to blush. That behaviour *could* be the basis of a particular belief, but if nobody ascribed that belief to the blushing person based on her behaviour and she never displays that particular belief, then that belief would remain entirely virtual, existing only as core elements. In this case, the core elements are causing her behaviour even though a belief is implicit in the core elements. Expressing, articulating, or ascribing are the only ways to make this belief explicit even though the

basis of that belief—the mental representations forming the basis of the belief—can influence the blushing person's actions.

When a person is visually experiencing the world, the brain is gathering visual information about the world. To illustrate, when I look at a telephone, my brain may represent the phone using multiple representations (such as a white handset, a grey body with a screen, and a long white cord), or my brain may represent the telephone simply using a single representation. Dennett maintains that the latter style of representation is probably very unlikely:

> If you were to sit down and write out a list of a thousand or so of your paradigmatic beliefs, *all* of them could turn out to be virtual, only implicitly stored or represented, and what was explicitly stored would only be information…that was entirely unfamiliar. (Dennett 1987b, p. 56)

In this quote, the term 'implicit' has the same meaning as the term 'virtual'. When a subject expresses or articulates an attitude, that attitude is made explicit; the attitude is only *implicit* (or virtual) in the mental representations prior to their expression (or articulation).

That being said, I don't think there's any reason to suspect that a belief attitude cannot, in principle, be stored as a single representation. I'll use an analogy to illustrate this point. Using a word processor, one might want to *copy* or *cut* a chunk of text from one page to another. There are (at least) two options: first, you can select the text, click the right mouse button, click the *copy* or *cut* button, then right click on the location of the document you want to paste the text, and click the *paste* button. Second, you can select the text, hit *ctrl-c* to copy the text, and *ctrl-v* to paste. The keystrokes in the second option

are *macros*. *Macro* is a technical term in computer science and an easy way to understand the concept is that a macro is an instruction initiating a series of other instructions to perform a task. In the previous example, a single keystroke accomplished a task that without the macro usually requires several steps to accomplish.

Both of these options perform the same task, but they each perform the task differently. A macro is analogous to single representation attitudes because there is no principled reason for a single representation to be unable to represent the same thing that a *group* of other representations could also represent. A single representation can accomplish what a group can accomplish; in fact, it's probably more energy efficient for a brain to use single rather than multiple representations in many cases. I wish to claim, using a popular turn of phrase, that *a representation is a representation*, regardless of how that representation is caused. Just as a macro initiates a set of instructions using a single keystroke (a set of instructions that would take the user a lot longer to follow without the macro), there is no reason a single representation cannot perform the same tasks as multiple representations.

When *x* forms a belief, say a belief that *the Bigfoot lives in Saskatchewan*, that belief has an underlying representational basis. Within the representational content of *x*'s mind, there is a pattern between various representations that can be (for example) expressed as the belief involving the Bigfoot. These representations are the core elements of *x*'s belief.

The connection between a psychological attitude and its core elements is interesting, especially considering the notion that the elements, like beliefs, are causally linked to behaviour. Core elements are not themselves paradigmatic beliefs; the core

elements aren't "beliefs *par excellence*" (Dennett 1987b, p. 56). They're the building blocks of a belief.

If our beliefs were the result of different representations working together, then a finite number of representations can give rise to a staggering number of various combinations. Additionally, if representations are partitioned into separate groups, then an even more staggering number of combinations would be available. Here's an example of what I'm talking about. The belief that *Santa Claus is real* is different from the belief that *the Bigfoot is real*. Each belief contains a similar predicate, *is real*, but they differ in their subject. The core elements causing us to express a particular belief about something *being real* might be shared by each belief and *being real* is a shared feature. Whether the predicate *is real* exists as a single representation, or exists in the connections between multiple representations, the core elements of *is real* are a shared feature between both beliefs *the Bigfoot is real* and *Santa Claus is real*.

It doesn't matter for the present discussion *how* a brain *specifically* represents the world (I mean how the representations are physically implemented in the brain), what is important is that brains *do* represent the world, whether those representations are based on external (such as a visual representation of the environment) or internal features (such as other beliefs, experiences, or the regulation of bodily functions), or whether content is represented using a single representation or multiple representations. This capacity for representing the world is a consequence of our evolutionary design (our capabilities as evolutionarily designed systems were discussed in Chapter I of this project). An intentional system must represent the world because the system is designed to have certain capacities for navigating the world, and to act appropriately in light of those

capacities. If a system couldn't represent the world, that system would be able to respond only to immediate events. There would be no memory, no learning, no private attitudes.

The brain represents the world, and some of those representations cause the formation of our beliefs, desires, fears, and other psychological attitudes. The core elements of these attitudes are the representational basis of a subject's behaviour that provides a basis for intentional stance ascriptions. Behaviour—whether that behaviour is in the form of either speech or an action—is caused by events in the brain.

The brain's capabilities and representations are the basis of our psychological attitudes. Beliefs are based on a representation of either the external or internal environment. We can display these psychological attitudes in our own behaviour, and the intentional stance relies on this behaviour for its ascription of attitudes. Publicly expressing an attitude, and ascribing an attitude, both rely on representational content. However, there's a difference between an attitude that's been publicly expressed and an attitude that's been ascribed. A person adopting the intentional stance toward me doesn't have the same access to my representational content as I do. This difference will be investigated in the next section.

2.4. Expressing attitudes vs. ascribing attitudes

We express our own attitudes and ascribe attitudes to other people. When we use the intentional stance to ascribe a belief or a desire to another person, we're basing our ascription on that person's behaviour. We have access only to other people's behaviour. An intentional stance adopter has access to either what its subject displays or has displayed in the past. As a consequence, the intentional stance can access only its

subject's display, a display that by definition consists entirely of behaviour.

On the other hand, when we form an attitude, we have access to more than our behaviour. When I say that I believe the traffic light is red or that my toe hurts, I'm not basing these beliefs on my behaviour. The formation of psychological attitudes is based largely on representational content, and one doesn't have to display that content in behaviour to form the attitude.

In section 2.2, I argued that representational content serves as the basis of our attitudes. The content of a belief's core elements represents the content of that belief, and our expressions display that content for other people. A subject has a different relationship to her own representational content from the relationship another person has to her content. This relationship is evident in the case of our private articulations; If I'm jealous toward one of my colleagues, but never express that jealousy, then obviously I'm not basing that jealousy on an observation of my behaviour. I'm basing my articulation on private, internal features that are unavailable to other people.

If, for example, that jealousy *was* displayed, then it would be fair game for the intentional stance. Behaviour that can be interpreted and used as a basis for an ascription of jealousy, this is the subject matter of the intentional stance. But the intentional stance cannot access a private articulation; only a subject's display can be the basis of an attitude ascription. In the Martian illustration I referred to in Chapter I section 1.2, the stockbroker was ascribed a *desire to buy stocks* based on his behaviour. Experiences can also warrant a psychological ascription; both Tom and Boris in the Trafalgar Square example are ascribed belief F (*a Frenchman committed murder in Trafalgar Square*) because they each have experiences that can be explained by ascribing belief F to them.

Psychological ascriptions using the intentional stance rely on its subject's experiences expressed in behaviour.

But since many experiences are never displayed in behaviour, those experiences cannot be the basis of a belief ascription by another. Additionally, some attitudes an individual might be able to form wouldn't be attributed to that individual, because that attitude may not appear salient (I may privately articulate that my window blind has thirteen slats, or that there are fifty-two ceiling tiles in this room). The intentional stance is limited only to behaviour, but the attitudes we form aren't.

Both myself and another person (adopting the intentional stance) have access to my representational content, except that access is different for each of us. We each access the content by forming a belief, but the primary difference is that the content is *mine*, rather than someone else's. What I can express (or privately articulate) is only *my* content. For example, if I'm walking down the stairs and I think the bannister is too sticky for comfort, another person will only be able to ascribe to me a belief based on the bannister's stickiness if that experience is displayed in my behaviour. I'm the one feeling the sticky bannister, not the person ascribing me the belief. The mental representations caused by the stickiness are accessible to the other person only if those representations are displayed in behaviour

The simplest way to understand my point is by recognizing that my representational content is my *own*; other people relate to this content only when it's displayed in behaviour. My relationship to my own content is different from another person's relationship to my content. If the intentional stance could magically access my private articulation of stickiness, in this case, then it likely wouldn't be able to ascribe

that belief to me because *that belief wouldn't predict or explain my behaviour*. In Dennett's view, the psychological attitudes that an intentional system *has* are supposed to be the attitudes that end up predicting or explaining that system's behaviour. But many of the attitudes that we form don't predict our behaviours. Yet I really have that belief (and so would you if you touched the bannister!) even though the belief is only privately articulated. The core elements of that articulation are caused by my interaction with the bannister, namely, feeling the stickiness of the bannister.

Other people (using the intentional stance) can infer only what experiences I'm having, based on my sensory exposure and my expressions. But I don't have to make this inference: I'm the one seeing, acting, and talking about my experiences. In the next section, I'll investigate my relationship to my own psychological attitudes using Dennett's notion of a theorist's fiction.

2.5. Theorist's fiction

Our psychological attitudes may be based on representational content, but what kind of access do we have to that representational content? In other words, when I form a belief, am I accessing representational content? So far, in the view that I've presented, a particular attitude appears to access representational content directly: beliefs and desires are based on my representational content. Dennett would likely find this idea of direct access questionable; the content of our beliefs based on only how the world seems to us, rather than how the world actually is. If beliefs are based on representational content (as I've been arguing in this chapter), yet what we're believing is only the world as it seems, then our mental representations represent only the world as it seems.

Dennett might call our attitudes and their representational basis a *theorist's fiction*. A theorist's fiction compares our attitudes to a *text*, and that text composes something like a fictional world (Dennett 1991, p. 81). An interpreter of that text (ourselves or other people) can derive truths from that text (p. 81). As people who form beliefs and desires about the world, these attitudes constitute how the world seems to us (p. 128) even if the way the world seems is accurate (p. 81). Our reports—the attitudes we form that may be publicly expressed or privately articulated—are theoretical constructs about the world (p. 157). These attitudes are "exquisitely useful fictions" (p. 367). Even though our attitudes are fictions, they can usefully describe how we feel, what we believe, and what we desire (for example). Dennett claims that it's much easier to "calculate the behavior of systems using this principled fiction than it would be to descend to the grubby details" (p. 367). We've seen how much easier it is to use a system's beliefs and desires to account for that system's behaviour in the Martian illustration from Chapter I, section 1.2 (pp. 24-25) of this project when using the intentional stance. In Dennett's view, our attitudes compose something analogous to a fictional world created by our brains. The attitudes involve only how the world seems to us, not necessarily (the fiction might end up being true, after all) how the world actually is.

Seemings are not particular, special states; we don't have separate representations of how the world seems to us and how the world is. All of our representational content is a seeming. Another way to understand seemings is to treat seemings as *judgments* (Brook 2000, pp. 234-235). Our representational content is a judgment of whatever stimulus the content is supposed to represent. A simple way to put this point is that the brain's inputs

(whether those inputs are from external or internal sources) are judged by the brain, yielding the content that's represented. Since our attitudes are contentful, and that content is simply a judgment of the stimulus, we are forming judgments about the world.

I'll attempt to make this connection between our attitudes and a theorist's fiction clearer; say that *x* formed a belief that he saw a Bigfoot. If we were interpreting *x*'s belief as part of a theorist's fiction, then x's belief that something (seeing a Bigfoot) happened to him (*x*'s self) is part of this theorist's fiction. Dennett maintains that *a self is just the centre of narrative gravity* (Dennett 1991, p. 431): *x*'s belief posits a self and ascribes particular experiences and attitudes to that posited self (1991, p. 418; p. 430). In other words, my brain posits a self (me) and an experience happening to that self. In Dennett's view, the content of our beliefs (as an example attitude) involve not only what's believed, but also who is experiencing that belief. A self is part of a theorist's fiction, and the entirety of our mental life constitutes a fictional account of what we judge to be happening. When we form our attitudes, we're forming judgments about what we seem to be experiencing, in Dennett's view.

When I say that *I believe D is a bigoted and crooked political candidate*, I'm saying only how the content of the belief seems to me. I'm judging D to be crooked and bigoted, and my belief is based on that judgment. My belief that D is crooked is a judgment based on that attitude's core elements. For Dennett, this judgment may or may not reflect how things actually are. The content of the belief itself, however, is part of my *fictional world*. The representational content my attitudes are based on is really a theory about my experience, how things seem to be. In Dennett's view, that representational content is part of my fictional world, that is, part of the world as it seems to me.

I agree with Dennett's view, partly. I agree with the *theorizing* aspect, but not the *fictional* aspect. I accept his view that the judgments generated by our brain can be analogous to a theory about how the world is, but our attitudes resulting from this theory are not adequately captured by the term 'fiction'. The reason for the term 'fiction' is that interpreting a person's attitudes and judgments is supposed to be analogous to interpreting a work of fiction. I think that instead of creating a fictional world with our judgments, we're doing something more like offering an inference to the best explanation. The reason we're expressing a fictional world, in Dennett's view, is because we're the apparent authors of that fictional world and "what the author (the apparent author) says goes" (Dennett 1991, p. 81). Just like the author of a text, as authors of our attitudes, we don't know why we say what we say, we only know what we say. We don't know how an author comes to know a certain aspect of his text (how does Doyle know the colour of Holmes' chair?) we simply take the author at his word (Dennett 1991, p. 81).

In the same way, we take our attitudes at face value: we know what we believe and desire, but we're not an authority regarding why we formed those attitudes. There may be elements of this view that are true, for example, there might be unconscious motivators causing us to form a particular attitude. But I think the view that our attitudes constitute a fictional world fails, and this failure can be seen when we consider how our experience is interpreted using the intentional stance. On the subject of the differences between interpreting texts and interpreting people, Andrew Brook writes in his article "Explanation in the Hermeneutic Science" (1995) that the difference between interpreting texts and interpreting people is a difference of interpreting two kinds of meaning; the first

kind of meaning involves how the different words and sentences relate to each other; the second kind of meaning—the kind that concerns interpreting people—involves how we're interpreting the meaning of attitudes by looking at their causes (Brook 1995, p. 11).

Since our attitudes are based on representational content, and that representational content is the result of our brain judging various stimuli, our attitudes are also based on judgments. I've claimed throughout this project that as judgments, our representational content is mostly accurate, that is, our representations accurately reflect what's being represented. So when we form an attitude, the content of that attitude accurately reflects its cause (in most cases). When I form a belief that D is crooked, whatever caused the core elements of that belief accurately reflect the world. Something real caused the core elements of that belief. In other words, the explanation of the causes of the core elements of my belief that D is crooked is best expressed as that particular belief, rather than any other belief. Explaining a particular attitude using causes is different from explaining how different aspects of a text relate to each other.

In Dennett's view, when a subject forms a belief, her knowledge of *why* she formed that particular belief is not supposed to be authoritative. When we state the causes of a particular attitude, we're doing so without any kind of authority. "You are not authoritative about what is happening to you, only what seems to be happening in you" (Dennett 1991, p. 96). But our brain's judgments—how the world seems to us—are mostly accurate. Our judgments are usually accurate—and thus the representational content forming the basis of our attitudes usually reflects the world—so our beliefs about the causes of our attitudes are also mostly accurate.

A subject has direct access only to its attitudes, not its representational content.

When I believe there is a carton of milk in front of me, I have access to my visual experience of the carton, and my attitude involving the carton. There is no carton of milk in my brain, and the attitude is explicit only in my articulation of the attitude. The representational content that my belief is based on contains that belief only implicitly; an attitude is made explicit when it's articulated (or put into words, whether those words are made public or kept private). The content of the articulated attitude is only the judgment of a stimulus. Since my belief is only a judgment about what I'm experiencing, I'm not an authority about my experience. Neither is my brain. I'm an authority about only what I *seem* to be experiencing. The reasons for my belief about the milk carton aren't necessarily the *real* reasons because I have direct access only to the attitude, not the representational cause of the attitude. Using Dennett's terminology, that subject isn't an authority on the particular *cause* of that belief. I agree with this point. A subject is only authoritative about how the world *seems* to her even though how the world seems usually matches with the world as it is. In other words, we're only an authority about our attitudes, not the representational content underlying those attitudes.

We can access our representational content only *indirectly* through our attitudes. Thus, even though representational content is the basis of our attitudes, when we self-ascribe an attitude we're accessing representational content indirectly through our other attitudes. This is a process of introspection.[24] Introspection requires that the subject *look inward* and attempt to find the reasons for an attitude. I have already claimed earlier in this section that this process is nothing more than a process of using judgments to explain other judgments—attitudes, in general, are judgments about our experience, and reasons

[24] See Chapter II, section 2.2 (p. 57n21) for the difference between *introspection* and *interoception*.

are themselves attitudes—and the status of each of these judgments is the same: they're all examples of how things seem. This network of judgments motivates Dennett's use of the term 'theory'. By forming judgments of other judgments, we're creating something like a theory about what we're experiencing.

There is a set of beliefs and desires that an individual can form that don't have to be the result of introspection: sensory beliefs, emotions, and pain often don't result from introspection. This creates a distinction between attitudes resulting from a process of introspection and attitudes that don't result from introspection.

Attitudes such as my belief that *there is a table in front of me* fall into the latter category. There is no introspection involved with this belief; I'm just judging a visual experience. An example of a belief falling into the former category is a belief such as that *I'm a shy, anxious person*. This belief may be the result of examining my past experiences and my reactions to those experiences: an examination like this is a paradigm case of introspection. I'm evaluating different beliefs to arrive at another belief. I'm forming a belief that *I'm a shy and anxious person* based on my other beliefs. A red wine drinker who adamantly proclaims that she loathes red wine, despite drinking red wine whenever she has the opportunity illustrates the same thing. She (the wine drinker) offers reasons to justify her drinking of the red wine, and those reasons are the result of introspection just like the reasons offered by the shy and anxious person. Both of these cases are examples of beliefs that can be the result of introspection.

That being said, if our representational content constitutes a *theory* of what we're currently experiencing, then that theory is evaluable as any other theory. The theory may be true or false. I might actually be shy and anxious, or I might not be. I can accept that

the content we're expressing constitutes a theory as long as the theory is neutral regarding whether the content of an attitude expression is accurate or not. I've been claiming that our judgments (as well as the attitudes based on those judgments) are mostly accurate because we're evolutionarily designed to form attitudes that reflect how things are. Some of our judgments may not be accurate, of course, due to introspection, bias, or bad information (when our senses fail, for example). But in most cases, our judgments *are* accurate because the source of the judgments (the representational content) is accurate.

Dennett's theorist's fiction isn't an entirely adequate account of our relationship to either our attitudes or our representational content. His view does have something to offer; namely, that the world seems a certain way to us and the content of our attitudes is based on something like a theory about the world as it is. As I argued in Chapter I, I think that our attitudes are mostly accurate because we have the capacity for forming attitudes that accurately reflect the world. If this were the case, then only *some* of the contents of psychological attitudes would be adequately describable as fictional. Evolution would have designed us to respond to the *world*, not with a fictional account of the world. The aspect of *theorizing*, I think, applies to our attitudes, because our brains are *idealizing* the world to ourselves, not generating a perfect replica. Our representations are mostly accurate, following our evolutionary design, but the notion that our attitudes constitute a fictional account of both ourselves and the world we inhabit is just plain wrong.

2.6. Conclusion

The attitudes we form are based on representational content. There are many different sources of this representational content: from our senses (such as our visual or

auditory faculties), from our memories, and from the other various interocepted content (such as pains or other feelings), to name a few examples. Our senses were designed by evolution to gather information about not only the world but also ourselves as accurately as possible; there is good reason to suspect that much of the representational content in our brains accurately reflects what that content attempts to represent.

Representational content is the basis of our attitudes. The intentional stance can use these attitudes when they're displayed for its predictions, but this access is more indirect than the access an individual has to her own representational content. This point is evident when we consider our privately articulated attitudes; private attitudes don't affect behaviour and are thus unavailable to the intentional stance. Yet we can and do privately articulate our attitudes. Even if those private attitudes *do* end up affecting behaviour, the intentional stance might not be able to explain that behaviour using just the attitude that we display. I might behave differently when I feel that the bannister is sticky, but not necessarily in a way that indicates that I think the bannister is sticky. An intentional stance adopter wouldn't be able to ascribe this attitude to me unless she knew that the bannister was sticky. My access to the core elements of my private articulation *this bannister is sticky* is different from the intentional stance's access.

My access to my representational content is different from another person's access to my representational content. For instance, the representations of stickiness in my brain constitute a judgment, and that judgment constitutes, in part, the core elements forming the basis of my belief *the bannister is sticky*. This judgment may be part of a theory about what's happening to me, but since my representational content is generally accurate, that theory is accurate more often than not.

We access our representational content indirectly via our attitudes; when we express one of our attitudes, we're expressing the judgments that our brain uses to represent our inner or external environment. In that way, we're theorizing about what we're experiencing, but that theory is most often accurate and doesn't describe experiences that can be adequately described as fictional.

In the next chapter, I will investigate the nature of the attitudes we form from our representational content. I will argue that our attitudes are idealizations of representational content using a process of abstraction.

Chapter III.

Idealizing Content

3.1. Introduction

In IST, psychological attitudes are both *abstractions* and *idealizations* of patterns in behaviour.[25] A belief ascription based on an intentional system's behaviour is an abstraction because ascribing that belief will provide a faster way to understand the system's behaviour than using all the data that can be gathered about the system. Remember Dennett's Martian illustration I quoted in Chapter I of this project? The stockbroker's behaviour can be described by either producing a report of all his individual motions and movements (such as raising his hand, placing it on the telephone, closing his fingers around the receiver, and so on) or by ascribing a *desire to buy stocks* based on his behaviour. This attitude is abstracted from the stockbroker's behaviour.

An attitude is also an idealization of a pattern in behaviour: the attitude is the one the subject ought to have, given his other beliefs and desires. The stockbroker is rational (and thus has the attitudes he ought to have), and given his other psychological attitudes (he wants more money, he believes that he needs to make a call to order more stocks, and so on), the stockbroker's behaviour can be understood by ascribing a *desire to buy stocks* based on that behaviour. Ideally, given his other attitudes and behaviours, a *desire to buy stocks* is the attitude the stockbroker ought to have. Other attitudes might be able to explain his behaviour, such as *a desire to buy GM stocks* or *a desire to own more shares*

[25] The idealized nature of beliefs and desires in IST was discussed in the Introduction of this project in section 3 (pp. 9-10). In this chapter I'll briefly reintroduce *idealization* in IST before I adapt the practice of idealization toward the brain's representational content.

in a company and thus more controlling power. All of these attitudes can help predict or explain the stockbroker's behaviour, but the *desire to buy stocks* is the simplest way to understand his behaviour. To summarize, a psychological attitude is both abstract and idealized; IST abstracts away from all of the messy details and idealizes his behaviour with the attitude *desire to buy stocks*.

Behaviour is a critical component of an intentional stance ascription. But not all of the attitudes we form are based on observing our behaviour. Often, we just say what we believe or desire without considering our present behaviour. For instance, I may privately articulate a feeling of dread, or a desire to go to bed early, but those articulations aren't based on me observing my behaviour. In this chapter, I'll argue that many of the psychological attitudes we form are idealizations of representational content. The core elements of our psychological attitudes,[26] in the view that I will present in this chapter, are abstracted from representational content and subsequently idealized using a particular attitude, an attitude that we ought to have given the rest of our representational content, beliefs, desires, and other attitudes. This will be made clearer in the chapter that follows.

In my view, our attitudes idealize representational content; this content consists of the brain's representations, from the representation of sensory or perceptual information to other attitudes. By adapting IST's notion of idealized, abstract attitudes to fit representational content instead of behaviour, I'm providing a major supplement to IST's analysis of intentionality, so that IST can account for the source of the attitudes that we form.

[26] See Chapter II, section 2.3 for an explanation of the core elements.

3.2. Two basic psychological attitudes

In IST, the two most important psychological attitudes are *belief* and *desire*. Intentional systems are rational systems, that is, they attempt to satisfy their goals (desires) using the information they think they have about the world (beliefs) (Dennett 1978, p. 6). These two attitudes are the most basic attitudes for predicting and explaining behaviour.

We form beliefs and desires as often as we ascribe them to others, so a theory of psychological attitudes must provide an explanation of the beliefs and desires that we both form ourselves *and* ascribe to others. IST relies on intentional systems displaying their attitudes, as I argued in Chapter I of this project. The intentional stance requires that a subject displays its mental contents in behaviour. This behaviour consists of, for example, the subject's actions and speech, but Dennett doesn't provide an account of the sources for the attitudes we form ourselves.

Dennett's explanation of the intentional stance most often involves the strategy's ascription of belief and desire based on behaviour, so I will focus on these two attitudes. A subsection will be devoted to each term. I focus mainly on 'belief', for two reasons. First, I'm using the term 'belief' very inclusively (any information that a system may possess can be considered a belief), and second, IST has more to say about belief than desire. Section 3.3 examines the role of belief in IST. I'll adapt IST's analysis of belief to accommodate how we form our beliefs based on *representational content* rather than an *observation of behaviour*.

In section 3.4 I will argue that beliefs are distinguishable as either an inferential

belief or a non-inferential belief. As I will use the terms, a belief that's *inferential* is inferred mostly from other beliefs (and other attitudes); a *non-inferential* belief is formed from all *other* sources excluding other beliefs and attitudes, such as sensory history.[27] I will provide a more detailed definition of both inferential and non-inferential beliefs in section 3.4. This distinction is a result of the representational basis of belief and serves to introduce the discussion in the following section. Section 3.5 looks at a potential problem with my view; I have been arguing (in Chapters I and II) that our attitudes are mostly accurate, that is, our attitudes are a reflection of not only how things seem to us, but also how things are. Confabulated beliefs are a possible counterexample to my view. I'll dismiss confabulation as a counterexample by providing an account of how confabulated beliefs could arise using the distinction between non-inferential and inferential beliefs from section 3.4.

Section 3.6 investigates the attitude of desire. In that section, I will provide an overview of how desires are ascribed in IST before I provide an account of desire attitudes based on mental representations. Section 3.7 examines the concept of *direction of fit*. Direction of fit is the feature that ultimately distinguishes beliefs from desires.

3.3. Preliminary remarks on 'belief'

Let's examine how IST uses the term 'belief'. I will adopt this account of belief in my supplement to IST. Most generally, beliefs are ascribed to systems based on patterns in that system's behaviour. However, what is it about any particular pattern that warrants

[27] My use of these terms is radically different from the standard use of the terms 'inferential' and 'non-inferential' in philosophy (see Chalmers 2014 for an example of the standard use of the terms). It's best to shed yourself of the meanings that the terms usually carry; in the context of this project I am using each terms *only* as I have described.

a *belief* ascription? The answer is very simple. When a system can be interpreted as possessing certain information about the world, the system can be ascribed a belief based on that information. That believed content causes the system's behaviour. For instance, *x* believes the door is closed, *x* opens the door before he walks through. The information contained in the belief *the door is closed* allows *x* to open the door before walking through it. He needs to see the closed door before he can open it. The intentional stance attributes that belief to *x* because the content of *x's* belief is caused by sensory information. We can surmise that *x* has this sensory information because that information is displayed in *x*'s behaviour of opening the door.

To summarize, IST claims that patterns in a subject's behaviour are the basis of belief ascriptions, which means that particular subject possesses certain information about the world. In the case of *x*'s belief *the door is closed*, *x* is ascribed a belief because *x* can be understood to possess that information. However, this specific explanation of the term 'belief' within IST differs wildly from the way ordinary conversation uses the term. It's important to recognize this difference; ordinary conversation is a little faster and looser with the term, whereas IST adopts a more technical approach to the term 'belief'. I will be relying on IST's use of the term 'belief' in this chapter.

Here is an example of this difference. In ordinary, everyday conversation, *x* likely wouldn't be described as *believing* that the door is closed, we would usually say that *x knows* that the door is closed, or refrain from saying anything at all. Dennett notes that in ordinary conversation,

> 'belief' is ordinarily reserved for more dignified contents, such as religious belief, political belief, or—sliding back to more quotidian issues—specific conjectures or

hypotheses considered. (1998b, p. 324)

Usually, we don't ascribe a belief based on mundane perceptions—such as seeing a closed door—to people; we usually apply that term to an individual's speculations about the world. In contrast to IST, ordinary uses, such as *believes that y*, are reserved for matters that don't involve sense experience. Whereas the intentional stance would attribute a *belief that y* based on the individual's behaviour, ordinary conversation might attribute something like *knowing that y*. Ordinarily, *knowing* something implies more certainty, or carries more significance, than *believing* the same thing. Under this use, *knowing that y* is supposed to imply that *y* is true and *believing that y,* in contrast, doesn't necessarily commit that *y* is true. In many cases, *believing* can serve as a hedge against falsity, or indicate a special kind of knowing, such as the "knowledge" that some religious people identify with their faith.

The distinctions between *knowing that y* and *believing that y* aren't relevant to the present discussion because these distinctions are not built into IST. Ordinary language's use of *to know* and *to believe* are counted under the rubric of *belief* in IST. I mention this distinction because I want to emphasize the difference between how beliefs are ascribed in IST versus the application of the term 'belief' in everyday conversation. There are similarities, however, between 'belief' in ordinary conversation and 'belief' in IST. In both cases, the ascription is based on the subject possessing certain information about the world (gained through sensory history in the case of *x*'s closed door). Remember Dennett's example of Sherlock, Jacques, Tom, and Boris (1987b, p. 54-55) that I discussed in Chapter I? Each person can be ascribed the belief *a Frenchman committed murder in Trafalgar Square* (F) despite their differing experiences. In ordinary conversation, we might say Sherlock and Jacques *know* that F (because Sherlock and Jacques were present for the

murder) and Tom and Boris only *believe that* F. Each case shows something similar: they all possess certain information regardless of how they each acquired that information, whether, as ordinary language claims, they *know* that F, or as IST claims, they *believe that* F.

The Trafalgar Square example illustrates how IST considers a belief to be an *idealization* of a pattern used to predict or explain behaviour. The belief is an idealization because it's the belief a system *ought* to have, and the belief coheres with the rest of the system's representational content. Tom and Boris, for example, are rational, intentional systems, and they ought to believe that F because they each read a newspaper containing a story about the murder. Since we know they both read a newspaper with the story about the murder in Trafalgar Square, the same belief is ascribable based on both Tom and Boris' experience. Not only is F a belief they all ought to have, but F is also a much simpler description of the information that both Tom and Boris likely possess than the information F is abstracted from. In other words, F is an idealization of information that Tom and Boris ought to possess, information that's abstracted from all of the separate bits of information they possess. The article on the murder didn't just read *a Frenchman committed murder in Trafalgar Square*; it would have been a much longer story with a lot more information, such as the name of the murderer, who apprehended the murderer, the time of day, and so on. Given that they're all rational, intentional systems, this abstracted information can be idealized as a belief that F. F is an idealized version of the information that both Tom and Boris should have, given that they both read a newspaper with an article about F.

However, there are some trade-offs with this practice of idealization. By

abstracting away from the details and using an idealized attitude, we're losing precision in our ability to predict behaviour using the intentional stance. We might be able to predict Tom and Boris' behaviour in some situations using F, such as if we were to ask them a question about whether there was a murder in Trafalgar Square last night. If we were to ask Tom and Boris about the details of the murder (for example), we couldn't predict their responses using F. Our ascription of F would fail to successfully predict or explain their behaviour because neither Tom nor Boris acquired any details regarding the murder that can be abstracted and idealized using F. Idealizing behaviour sacrifices precision but in most cases idealizing behaviour works just fine.

That being said, not all beliefs are idealizations *of behaviour*. What if Tom or Boris said to himself privately, *oh it looks as if a Frenchman murdered someone last night*. They're not idealizing their own behaviour in this case. Their articulation is an abstraction from the information they've gained from reading the newspaper, not from behaviour. This information is idealized as the belief that F. I'm focusing on the representational basis of such psychological attitudes, as a supplement to the account of attitudes in IST. Claiming that representational content, is a basis of psychological attitudes, in along with behaviour, constitutes a major addition to IST.

In the preceding section, I outlined how beliefs are abstractions and idealizations in IST. Now that I've provided this outline, I will begin adapting IST's treatment of belief to allow for representational content, rather than patterns in behaviour, to act as a basis for the beliefs we form. In light of this outline, in the next section, I'll argue that an intentional system's capacity to represent the world—bestowed on the system by being susceptible to the intentional stance—means that there is a distinction between two types

of belief that isn't obvious when one looks only at behaviour. I will refer to these two types of belief as either *inferential* or *non-inferential.* These terms will be explained in the next section.

3.4. Two types of belief

Beliefs can be inferred from other beliefs. What I mean is that the content of a belief can depend on other beliefs. If I believe it's still raining in Ottawa, for example, and I also believe it would be wise to bring along an umbrella when I travel into the city so I can stay dry, then my belief involving the umbrella is inferred from my belief that it's raining in Ottawa. My belief that it would be wise to bring an umbrella with me to Ottawa requires that I also believe that it's raining or that it's going to rain in Ottawa. Another example of a belief inferred from other beliefs is the case of someone who believes that the Bigfoot lives in the forests of Saskatchewan. That person's belief is inferred from whatever information the person thinks supports his belief that *the Bigfoot lives in Saskatchewan*.

So far in this project, I have been considering mostly sensory beliefs. However, the intentional stance ascribes beliefs to an intentional system based on more than just the system's sensory experiences. For example, Art may attribute belief *p* to Betty because Betty also has beliefs *q* and *r*. Belief *p* can be inferred from *q* and *r* together. In this case, Betty has *p* because of her other beliefs. In this section, I will argue for a distinction between *inferential beliefs* and *non-inferential beliefs*; beliefs that are inferred from other beliefs are *inferential beliefs*, and beliefs acquired from all other sources excluding inferences from other beliefs are *non-inferential beliefs*. I will elucidate this distinction in

the following section. In Chapter I of this project, I argued that intentional systems have certain capabilities that are a consequence of being both rational and evolutionarily designed. We have the capacity to form beliefs based on current sensory information, but we also have the capacity to form beliefs based on *past* sensory information. In addition to sensory beliefs and memories, we also form beliefs involving states such as pains or emotions. When I stub my toe and articulate that pain, the articulation isn't based on behaviour. When I'm feeling a sense of dread, and I articulate that feeling, my articulation isn't based on behaviour. In other words, we have the capacity to publicly express or privately articulate content that's represented in our brains using abstract psychological attitudes. For example, that content may be the result of sensory beliefs, memories, or feelings (such as disgust, pain, or thirst for example), all of which may provide the basis for a psychological attitude. How do these capacities impact our psychological attitudes?

Consider these two examples:

(1) I believe that there's a telephone on the desk

(2) I believe that the Bigfoot lives in Saskatchewan

The first example illustrates the kind of belief that's been primarily discussed so far: a belief that's based on sensory or perceptual experience. These kinds of information (sensory or perceptual experience) are one major source of belief production. Someone asks me where the telephone is; I respond by telling them that the telephone is on the desk because I can see it there. The core elements of this belief consist of the various mechanisms and representations caused by the brain's inputs, which ultimately cause the belief expression. The example in (1) serves my purposes just fine, but we can also recast

(1) into something such as:

(1a) I believe that my toe is broken

I express this belief based on the consequences of smashing my toe against the wall, not because I'm making an inference regarding my behaviour involved with smashing my toe against the wall. In both (1) and (1a), ascribing this belief to myself isn't based on behaviour. Both (1) and (1a) are based on my experience, and the core elements of each are caused by the way my brain and body interact with the world (through vision, for example) to produce representations based on my experiences (either visual or nociceptive representations, in these cases).

The second example may appear to be the same because (1) and (2) are both beliefs. However, the content of these beliefs must have a very different basis, and not just because the content of each is different. If someone were to ask *x* where the Bigfoot lives and he responded with (2), then the core elements of the belief cannot consist of any sensory representations, whereas the core elements of (1) are sensory representations. *X* wasn't visually confronted with the Bigfoot in the Saskatchewan wilderness (nobody has seen the Bigfoot unless you think the Bigfoot is only a guy in a furry suit). By *x*'s own admission, he's never been to Saskatchewan, so the basis of his belief cannot be memories of a previous sensory event. It's possible that *x heard* that the Bigfoot lives in Saskatchewan, but if so, *x*'s belief would similar to (1), in that *x's* belief is based on his *hearing* about the Bigfoot's residence in a way that could also lead to an observer ascribing (2) based on *x*'s auditory experience. In other words, some versions of (2) could be formed based on sensory information, but others aren't. The core elements forming the basis of a non-sensory belief must be different from the source of beliefs directly linked

to experience.

(1), (1a), and (2) are examples of beliefs that might be expressed in ordinary conversation. Most of the beliefs that we express are like the beliefs in the second example, (2). Political beliefs, cultural beliefs, religious beliefs—none of these types of belief are directly based on sensory history. Instead, beliefs such as (2) are inferred from other beliefs, similar to how Art can ascribe *p* to Betty based on Betty's other beliefs. Most of the beliefs expressed in ordinary conversation are a lot more complex than the beliefs ascribed by the intentional stance, and require the support and existence of other beliefs, such as the aforementioned political and religious beliefs. (2) is based on other beliefs comparable to the way that beliefs from the first example are based on experience. Beliefs can be inferred from other beliefs, and this particular principle is the second major source of belief generation. The core elements of this type of belief, illustrated in (2)—which I'll call *inferential beliefs*—are not caused by any immediate sensory experience.

Beliefs such as the one illustrated in (2) are *beliefs inferred from other beliefs*. The core elements of a belief such as (2) may involve other beliefs, such as the Bigfoot's migratory habits, the Bigfoot's environmental preferences, and so on. This is different from a belief such as the one illustrated in (1) because the core elements of (1) are directly related to experience. Dennett's very minimal definition of belief, *any information that causes behaviour* can be easily adapted to accommodate *beliefs inferred from other beliefs.* The content in beliefs such as (1) or (1a) finds its basis in experience and the content in beliefs such as (2) is based primarily on other beliefs through something like an inference. As I claimed above, it's possible *x heard* that the Bigfoot

lives in Saskatchewan, and if he forms this belief, then *x*'s belief would be more like (1) because his belief is based on his "hearing" about the Bigfoot's supposed residence. Beliefs such as (2) *can* be due to sensory experience, but most often, they're not. Beliefs requiring other beliefs (such as (2)) aren't based directly on experience, by definition. The core elements of an inferential belief may include sensory representations, but the inferential belief is an attitude inferred from predominately other attitudes.

One potential problem with the way I've described inferential beliefs so far is that if an inferential belief depends on other beliefs, then it may seem as if an inferential belief requires that all of our other beliefs are stored in our head in some way. To illustrate this problem, consider *z*'s belief that she's shy. For *z* to have a belief with content involving her shyness, she needs to have other beliefs, such as a belief about *what shyness is*, and beliefs about herself that could support her belief about being shy. *Z*'s belief that she's shy requires her to have other beliefs, beliefs that must be represented *somehow* in her brain. The problem with this approach was introduced, and rejected, in Chapter II, section 2.3 (pp. 62-65) of this project. The vast number of beliefs inferrable from *z*'s original belief poses a problem of how all those beliefs are stored. The existence of beliefs as explicit brain states, representations whose content is identical to the expressed content, was rejected in Chapter II, section 2.3, because of the explicit storage of separate beliefs very quickly leads to a problem of storage space.

Instead, inferential beliefs are like non-inferential beliefs—abstract and idealized. (1) and (1a) are examples of non-inferential beliefs; neither of these attitudes are inferred from other attitudes. Treating beliefs generally as abstract, virtual states rather than concrete, explicit states can circumvent the storage problem faced by concrete treatments

of belief. For the case of inferential beliefs, I'm claiming that the source of the core elements of inferential beliefs are abstracted from other beliefs, rather than sensory information (for example). So, when *z* claims that she's shy, that belief is inferred from her other beliefs. Her belief that she's shy is abstracted from other beliefs. When *z* forms the belief *I am shy*, she's abstracting that from the other beliefs forming its basis, and idealizing that abstraction with the belief *I am shy*. This example illustrates how the attitudes we form are idealizations of abstracted content. *Z*'s belief that she's shy is an idealization of her other beliefs because it's the belief that she ought to have given her other beliefs. The core elements of *z*'s belief about her shyness are abstracted from the rest of her representational content, and idealized as the belief that she's shy.

Beliefs are abstracted and idealized using an inferential belief. An inferential belief, such as *z*'s belief about her shyness, is an idealization of other attitudes. Inferential beliefs are abstracted from other attitudes and idealized using a particular attitude depending on how the abstraction coheres with the rest of the individual's attitudes. Remember how Tom and Boris's non-inferential belief that F from section 3.3 of this chapter was an idealization of the information they were exposed to via reading the newspaper article about the murder? *Z*'s belief about her shyness is also an idealization of other information, in this case, an idealization that she ought to believe, given her other beliefs. The core elements of her shyness belief are abstracted from her other beliefs (and other representational content) and idealized using the inferential belief *that she's shy*.

Inferential beliefs (such as in (2)) are idealizations of other beliefs. Beliefs (non-inferential beliefs such as in (1) or (1a)) are also idealizations of core elements that have been abstracted from other representational content. When I express a belief about visual

experiences—such as the telephone in front of me— the core elements of that non-inferential belief are abstracted from the details of my visual experience (such as the phone's colour, shape, texture, location, etc.), and formed using an idealized attitude. The belief is idealized because it's what I ought to believe, abstracted away from the details. It's a belief that I wouldn't know that I had until after its core elements were idealized. Similarly, my expression about my broken toe is an idealization of other various aspects of my experience. Rather than saying something like *my toe has a dull throbbing pain* and *my toe is black and blue,* I'm forming a belief by abstracting away from the details and idealizing the abstraction as *my toe is broken*.

To summarize: our beliefs can be divided into *non-inferential beliefs* or *inferential beliefs*. A regular non-inferential belief is an idealization of representational content. In other words, when we form beliefs about features of our experience (such as what we're seeing, hearing, or feeling) we're forming a non-inferential, idealized version of what we're experiencing. A belief such as *seeing a telephone* is an idealized, non-inferential belief because the belief is summarizing features in the environment. In the case of this non-inferential belief, one is seeing a lot more than just a telephone, even though only a belief about a telephone is formed. An inferential belief, on the other hand, is an idealization of other beliefs. The core elements of these other beliefs are interpreted as a new, different inferential belief. Believing that you're shy, or that the Bigfoot lives in Saskatchewan, are both examples of inferential beliefs whose core elements serve as an idealization of other beliefs.

Many of our beliefs (whether they're inferential or non-inferential) are not based on a single source, such as our senses, memories, other beliefs, or even behaviour all by

itself. I have focused on the fact that representational content—internal features—can serve as the basis of our beliefs in addition to behaviour. The basis of our beliefs is more complicated than it first appears.

I have claimed in Chapters I and II that our beliefs are mostly true, meaning that we're not only expressing how things seem to us but also how things are. How things seem to us is most often how things are. That being said, there is an important phenomenon suggesting that our beliefs don't always reflect how things are, only how things seem to us. This phenomenon is called *confabulation*, and it will be the subject of the next section.

3.5. Confabulation

In the previous section, I distinguished between non-inferential and inferential beliefs. I described inferential beliefs as beliefs inferred from other beliefs (or other attitudes generally). How many "orders" or how complicated a particular inferential belief's pattern is doesn't matter; an inferential belief is an idealization of a pattern in other attitudes.

I also claimed that mostly, our beliefs accurately reflect the world. What about cases where we're mistaken? Being mistaken about something, whether the mistaken belief is inferential or non-inferential: sometimes our senses falter, or we receive faulty information. In some rare situations, however, neither of those two things happen, and when prompted to provide a reason for our expression we *confabulate* the reason. A person is *confabulating* when he "makes a false claim without an intent to deceive" (Hirstein 2005, p. 3). A confabulated belief is a false claim, a claim where the person

making the claim thinks the claim is true. Confabulation isn't lying because while lying is purposely making a false claim, a person expressing a confabulated belief doesn't know her belief is false. A confabulated belief isn't based on faulty information; confabulated beliefs are entirely made-up.

Here's a classic example of confabulation from Nisbett & Wilson (1977). The researchers set up a table in a public space with pairs of stockings on a table. Passerbys were asked to rate which of the stockings they preferred. Different passerbys gave different answers; some claimed that one pair of stockings had a better texture, others claimed that they preferred the colour of one pair over another. However, each pair of stockings was identical—they didn't vary in size, texture, or colour. The passerbys each provide confabulated answers: they were expressing false beliefs that didn't reflect the world, without knowing that they were expressing false beliefs. All of the stockings had the same texture and colour, so their beliefs about the difference in colour or texture weren't based on experiencing differences in the stockings. Their preferences were confabulated.

I have claimed that our beliefs are generally true, that is, our beliefs both accurately reflect how things seem to us *and* how things are in the world (there's often little difference). However, as seen in Nisbett & Wilson's study, the passerby's beliefs were true only about how things seemed to them, not how things are in the world. The gap between *how things seem* and *how things are* is exemplified by confabulated beliefs.

Dennett discusses this gap between *how things seem* and *how things are* in *Consciousness Explained* (1991). The gap is supposed to be bridged using *heterophenomenology*, Dennett's method for studying consciousness. A subject is studied

heterophenomenologically by combining the methods and tools of physical science with that subject's reports about his experience. Using experimental data along with interpretations of the subject's reports, Dennett claims that the heterophenomenologist can figure out "what it is like to be that subject—in the subject's own words, given the best interpretations we can muster" (Dennett 1991, p. 98).

A heterophenomenologist, according to Dennett, should be able to reconcile how things seem to the subject with how things are for that subject. Why a particular passerby selects one pair of stockings over another, despite no physical difference, should be explainable by a heterophenomenologist by comparing the subject's reports to the experimental data. The human susceptibility to confabulate the reasons and causes for our beliefs is another example of the difference between how things seem to us and how things are. Nisbett & Wilson's passerbys illustrate this distinction nicely.

This gap between *how things seem* and *how things are* produces a difficulty for the view that I'm presenting. As I wrote earlier in this section, I have been claiming that our psychological attitudes are generally accurate with regard to both how things seem to us and how things are. If that's the case, then how can the red wine drinker confabulate her reasons for drinking red wine? Her reasons seem like her real reasons, but they're not the real causes of her behaviour. How can the passerbys in Nisbett & Wilson prefer one pair of stockings over the rest? One pair of stockings may seem as if it has a more pleasing texture or colour, but all of the stockings are the same.

The existence of confabulated beliefs creates a problem for self-ascriptions of belief. Because self-ascriptions of belief are interpretations, and some of the beliefs that are interpreted may be confabulated, these intentional stance self-interpretations will be

less accurate than the standard third-person interpretations of the intentional stance. An individual who is interpreting her own beliefs is engaging in an activity that's rife with more personal bias than interpreting another person's expressions. Because we're actively looking (in other words, introspecting) for the basis of our beliefs, we're more likely to be subject to implicit or unacknowledged biases which are distinct from the biases that another person might have ascribing the same belief based on our behaviour. The wine drinker might be able to interpret alcoholism as the motivator for her (the wine drinker's) belief, but that process is long and filled with more activity than our effortless ability to form beliefs. It's more likely, however, that she is motivated *not* to see the truth behind her red wine drinking. It can be difficult to admit when you have a drinking problem, and it's easier to avoid or ignore the problem than confront it directly. Our beliefs often aren't the result of active interpretation.

In the previous section (3.4) on inferential and non-inferential beliefs, I argued that both types of belief are idealized abstractions of other content. Non-inferential beliefs are idealizations of experience (for example), and non-inferential beliefs are idealizations of other psychological attitudes. If our beliefs (in general) are considered to be idealizations of other content, then confabulated beliefs are also idealizations. Confabulated beliefs are the result of biases or pressures forcing the confabulator to form an attitude *without core elements accurately reflecting how the world is*. In Nisbett & Wilson's study, a passerby doesn't have any content representing the differences in texture or colour, but the passerby still forms a belief about the difference in texture or colour. The pressures of the situation (e.g., they expect that there's a difference between the stockings because an authority figure—the researcher—is asking them a question

about the differences between the pairs of stockings) force a passerby to form a belief about differences that aren't there. The passerby's brain abstracts whatever information it can to produce a belief in response to the researcher's question.

I have argued that our beliefs (whether the belief is inferential or non-inferential) depict not only how the world seems to us but also how the world is. The existence of confabulated beliefs, however, implies that these confabulated beliefs depict only how things seem to us. The vast majority of our beliefs aren't confabulated, but the minority of beliefs that *are* confabulated must also be accounted for in my theory.

I think that the similarity between confabulated and unconfabulated beliefs holds the key to explaining how confabulated beliefs are possible. Our ability to form both confabulated and unconfabulated beliefs is a sign that the core elements of a confabulated belief are idealized in the same manner as the core elements of an unconfabulated belief. In both cases, the attitude is an idealization of core elements that have been abstracted from other representational content. Following this point, the main difference between confabulated and unconfabulated beliefs is the availability of representational content for idealization. If a subject is asked a question, and there is no representational content available to idealize, the brain finds other representational content to idealize instead.

Another, less technical way to put this is to say that if there aren't any representations to support your belief, make some representations up! Consider Nisbett & Wilson's study: the researchers are asking the passerbys about the stockings, and even though there are no differences between the stockings, the brain finds a way to believe that there are differences even though there isn't any representational content directly caused by the differences in the stockings (because the stockings are all the same). A

passerby's brain is producing beliefs about the stockings without any representational content directly caused by the stockings to base those beliefs on.

If inferential beliefs are considered, then any potential problems for my view caused by the existence of confabulated beliefs go away. First of all, confabulated beliefs are in the minority; the vast majority of our beliefs are accurate because the source of the information underlying the belief is accurate. Second, confabulated beliefs are idealizations of content, but the content being idealized isn't content that's caused by the event associated with the confabulated belief. Passerbys in Nisbett & Wilson are forming confabulated beliefs because there's no representational content caused by the stockings associated with the differences they express, even though the researchers are asking questions about differences between the stockings. There's no content caused by differences in texture or colour, and the lack of this content causes the passerby's brain to base the belief on other, unrelated content. The suggestions of the researchers force the passerby to abstract some core elements away from representational content that's not caused by the stockings.

3.6. Desires

What is the basis for the desires I form? Desires are similar to beliefs, with one major difference. Before I introduce the major difference, I'll outline how beliefs and desires are similar.

Beliefs and desires play equally important, yet different roles in folk psychology and IST analyzes these two basic intentional states. For the intentional stance, desire attribution works similarly to belief attribution. Subjects are attributed desires they

"ought to have", and the intentional stance is supposed to consider the subject's most basic needs or goals when making an ascription (Dennett 1987a, p. 20). A subject's desires are, broadly construed, the subject's *goals* and a subject's goals motivate that subject into action. Thirst (the biological need for hydration, in this case), for example, can be the basis for a desire for water, and that desire motivates the subject into action. She is motivated to find water to drink. Thirst is a very simple basis of a desire, as is a subject's other biological needs, such as food, water, and procreation. Dennett puts this point more bluntly: "Trivially, we have the rule: attribute desires for those things a system believes to be good for it" (1987a, p. 20). Intentional systems are motivated to pursue things that it believes are good for it, like food and water, and those things the system believes to be good for it are often the content of particular desires.

Of course, we desire a lot more than food, water, and procreation. Here's a simple example: I want a television. A television doesn't satisfy anything biological and has nothing to do with food, water, or procreation. The intentional stance could ascribe a desire for a television based on my behaviour, but the stance could only make that ascription if I displayed that desire. I would need to say that I want a television, or go to a television store, for the intentional stance to ascribe that particular desire, even though I can (and often do) readily and easily articulate that desire privately without going to a store or making my desire public. If I wasn't displaying this desire, then the intentional stance won't be able to ascribe this desire. There's no behaviour to base the desire ascription on! Before the intentional stance could ascribe a desire, I must be able to display my desire using some form of behaviour as a basis for the ascription.

How I display this particular desire for a television is a different story. My desire

for a television is formed using the same process as a belief. In the view that I've been presenting, a belief is an idealization of core elements that have been abstracted away from all the messy details. Here's an example to remind you of what I mean. *X* believes the Bigfoot lives in Saskatchewan. The representational basis of this belief consists of (for example) other attitudes and past experiences. This representational content can be abstracted to yield core elements that are idealized using the belief *the Bigfoot lives in Saskatchewan*. This is a belief he ought to have, given the other attitudes and experiences he possesses, that's been abstracted away from all of the messy or irrelevant details present in the representational content. The core elements resulting from this abstraction are idealized using this belief.

The desires we form are the result of a similar process. Here's an example. A stockbroker forms a desire to buy some stocks. We could offer other attitudes to explain *why* he wants to buy stocks, such as making money, owning more share in a company, gaining power in the business world, etc. There may even be unconscious or unacknowledged attitudes motivating his behaviour in addition to the motivations he might express or articulate. These reasons can be combined and abstracted to yield an idealized *desire to buy stocks*.

Our desires are idealizations of other representational content which have been abstracted away from all of the related messy or irrelevant details. I may privately articulate a desire for a television, but that desire is based on other attitudes. Examples of these other attitudes might be something such as *I need a way to relax,* or *I would like to watch movies on a screen larger than fifteen inches*. Abstracting away from these two example attitudes can yield the idealized desire *I want a television*. This particular desire

expresses the information contained in those attitudes in a much simpler way than either attitude alone: it's much simpler to say that I want a television than it is to list all of the reasons I want a television.

Desires are idealizations of representational content in a way similar to beliefs. One major difference between the two is that desires involve some form of satisfaction; when *x* desires *y*, *x* will usually attempt to change something about the world to satisfy this desire. This difference leads to the way each kind of attitude relates to the world, and will be the subject of the next section.

3.7. Direction of fit

Beliefs and desires influence each other; *x* desires *y* because *x* believes that *y* will be good for it. This influence is a consequence of being rational: an intentional system uses its beliefs to figure out how to satisfy the desires motivating the system. Even though the attitudes are related to each other, there is a distinct difference between them. The way an attitude relates to the world will determine whether the attitude is a belief or a desire.

Here is what I mean. A classic way to distinguish between attitudes such as belief and desire is to refer to each attitude's *direction of fit*. Direction of fit was first identified by G.E. Anscombe in her classic monograph *Intention* (1963), although she never used the term 'direction of fit'. Anscombe uses an example of a man with a grocery list going to the grocery store while being followed by a private detective who is making a record of the man's actions. If *what the man buys* does not agree with the *list* (perhaps he bought an item that wasn't on the list), then the mistake is with the man's actions (i.e., what the

man buys). If the detective's *record of the man's actions* doesn't agree with *what the man actually bought*, the mistake is in the detective's record (§32). What we have here are two possible places for a mistake: between the man's actions and his list, and between the detective's record and the man's actions.

What these two possible mistakes are supposed to illustrate is a mistake of fit, first between the man's actions and his list, and second, between the detective's record and the man's actions. These mistakes can be summed up as how the words on a page (either the list or the detective's record) fit with the world, and what has to change to correct this mistake. There are two distinct directions of fit, depending on what has to change, the man's actions or the words on a page. If the man's actions don't fit with the list, then the man's actions must change; if the detective's record doesn't fit with the man's actions, then the detective's record must change.

Direction of fit has been applied to the way that mental states (such as beliefs or desires) fit the world. John Searle has used the term 'direction of fit' to describe a feature of mental states, although the concepts associated with the term have been discussed by philosophers before Searle. I'll use Searle's terminology to describe direction of fit because, so far, I find his terminology to be the most vivid elucidation of the concept.[28]

Searle claims that beliefs have *mind-to-world* direction of fit; beliefs are supposed to reflect something true about the world. The content of a belief *fits* the world, in other words, beliefs attempt to "match reality" (Searle 2004, p. 118). Desires, on the other hand, are said to have *world-to-mind* direction of fit, because what makes an attitude a

[28] It has been pointed out to me that Searle uses this terminology backwards. This may be the case, but the proceeding discussion will use Searle's terminology anyway. I was first introduced to the terminology in the following section through Searle, and his usage has always stuck with me (for good or ill).

desire is that the attitude motivates its bearer to change the world to fit the content of the desire.

> But in the case of desire, it is not the aim of the desire to represent how things are but rather how we would like them to be. In the case of the desire it is, so to speak, the *responsibility of the world to fit the content of the desire.* (Searle 2004, p. 118)

Beliefs attempt to reflect the world and desires attempt to change the world, according to this view. In IST, the way both attitudes are described matches up with the direction of fit view. If an intentional system is said to have a particular belief, then it's understood that the belief is ascribed to that intentional system based on the information about the world the system possesses, e.g., a driver will stop at the traffic light because she sees that the light is red. Experience can inform beliefs reflecting the world or have a mind-to-world direction of fit. When a desire is ascribed based on an interpretation of an intentional system's behaviour, this type of attitude ascription suggests that system is motivated to pursue things in the world. To uses Searle's terms, a *desire* attitude has a world-to-mind direction of fit.

The difference between the direction of fit of belief and desire attitudes is the key difference between those two kinds of attitudes. A belief attitude is based on information possessed by the bearer of the belief, and a desire attitude is based on something the bearer of the desire wants to change about the world. How representational content is idealized will depend on the direction of fit belonging to the core elements that are abstracted from that content. The core elements a desire is based on representations of the desirer's motivation for satisfying a goal (such as the physiological processes involved

with dehydration) unlike the core elements of a belief, which represent motive-free information about the world (such as what's in your current visual field, rather than what you *want* to be in your visual field). Direction of fit displays an intentional system's ability to form two distinct types of attitudes.

3.8. Conclusion

Psychological attitudes, whether the attitude is a belief or a desire, are idealizations of underlying representational content. We have inferential and non-inferential beliefs, which are two different kinds of mind-to-world attitudes. Beliefs in general can be based on experience, or inferred from other attitudes. I can express or articulate a belief based on something that I see in the world, or based on an emotion, sensation (such as dread or pain), or inferred from other attitudes without referring to my behaviour at all.

Non-inferential beliefs are idealizations of representational content. When I say that I see a telephone, I actually see a lot more than just a telephone. An aspect of my experience is abstracted and then idealized using a belief. Forming a belief about my toe being broken can be explained the same way as the visual example; the core elements of that belief are abstracted from my representational content and subsequently idealized as the belief *my toe is broken*.

Inferential beliefs work the same way, except that an inferential belief idealizes other beliefs. The representational content that's idealized consists largely of other beliefs, and not sensory experience, so this raises the possibility of error. Non-inferential beliefs, as I have argued in Chapters I and II of this project are largely accurate; in other

words, non-inferential beliefs generally reflect how things are. Inferential beliefs are also generally accurate, but there's more room for error. Faulty information, biases, and confabulation can more easily damage the credibility of a non-inferential belief.

Desires, on the other hand, are world-to-mind attitudes: forming a desire means that you are motivated to pursue something, that you want to change something in the world to fit your desire. Similar to belief, desires are also idealizations of representational content. The core elements of a desire that we form are abstracted from representational content and idealized using a particular desire, similar to belief.

In this chapter, I argued that some of our beliefs and desires are idealizations not of behaviour, but of representational content. IST claims that psychological attitudes are idealizations of patterns in behaviour, but that theory fails to account for the attitudes we form in our everyday psychological life. I claim that representational content is the basis for some of the attitudes we form. By extending IST's analysis of intentionality to include representational content, I believe that I have found a way for the analysis of intentionality presented by IST to account for psychological attitudes that aren't formed based on an observation of behaviour.

Conclusion.

Dennett doesn't provide any real account of the sources of our psychological life. The psychological attitudes that we form ourselves don't succumb to the same analysis as the attitudes that we ascribe to others. The attitudes we ascribe to others are based on, mostly, their behaviour and the sensory experience that may be involved with that behaviour. In *Consciousness Explained*, Dennett offers an idea that might be the beginnings of an account of our psychological life, but he doesn't go down that road. More often, he claims that our psychological attitudes are the result of adopting the intentional stance toward ourselves, which means that we ascribe attitudes to ourselves based on interpreting our behaviour and experience.

Yet we very rarely do this; most of the attitudes that we express to other people we don't ascribe to ourselves, and the attitudes we keep private are often ones that wouldn't be ascribed to us. At the very least we don't base our psychological attitudes on an observation of our behaviour.

This is a gap in Dennett's theories; how do we form—or publicly express, or privately articulate—our psychological attitudes? We don't do it on the basis of observing our behaviour. I have argued in this project that the basis of our psychological attitudes is, in fact, representational content. We can adapt Dennett's analysis of psychological attitudes and apply that adaptation to our own representational content. This adaptation constitutes a supplement to IST and the intentional stance: instead of behaviour as the basis for our psychological attitudes, representational content is the basis for the attitudes we form, express, and privately articulate in our everyday life.

My supplement to Dennett's theories goes like this. Intentional systems have certain capabilities. The intentional stance requires that those capabilities are displayed in behaviour for the purpose of making successful predictions and explanations. That an intentional system has particular capabilities carries some accompanying consequences; if an intentional system is capable of holding memories, for example, and acting on those memories, then there must be some way to store and represent the experiences contained in those memories. If I remember an experience and act upon that experience (say, remembering not to put my hand on a hot burner) that particular memory affects my behaviour in a way that can be discerned using the intentional stance. A belief or a desire can be ascribed to me based on that behaviour. Yet what if I don't make that memory public by not letting it affect my behaviour? I can still form the attitude; it's just that the attitude doesn't affect my behaviour.

The fact that I can form attitudes and only privately articulate those attitudes means that I have capabilities that may be invisible to the intentional stance. It's possible that the attitude will affect my behaviour without me realizing it, but in many cases, I can keep that attitude completely private. I think that this shows that we're capable of representing the world and our experience. The representational content resulting from this particular capacity is the basis for our psychological attitudes.

In IST, attitudes are both abstract and idealized. I adopt this position of abstract idealization for the attitudes that we form, whether the attitude is expressed or privately articulated (for example). Instead of abstracting from and idealizing behaviour based on another person's sensory experience, representational content serves as the basis for the abstractions and idealizations that we ourselves form. Here's how it works. The core

elements of a particular belief, say I believe *that there's a telephone in front of me*, are abstracted from the representational content those core elements are a part of. My visual experience is a source of a great deal of content; the core elements of the belief are abstracted from that visual representational content. Those core elements are subsequently idealized using the belief *there's a telephone in front of me* because all of the details involved with the rest of my experience are left behind in the abstraction. The idealized belief, *there's a telephone in front of me*, is a belief that I ought to have, given my visual experience.

To summarize: if we examine the capabilities required by the intentional stance, we find that some of those capabilities involve representing aspects in our experience. Those representations form the basis of our psychological life. There is a great deal of representational content in our brains, and by abstracting away from all of the content's details and idealizing that abstraction using a psychological attitude, we can, for example, *express* that attitude to other people.

I have attempted to broadly maintain Dennett's analysis of psychological attitudes, but instead of relying on the behaviour and experience of other people as the basis for these attitudes, I attempt to base psychological attitudes in the representational content within our brains. The representational base of psychological attitudes serves as the source for the attitudes we ourselves form, express to others, and articulate privately. By basing our attitudes on representational content while maintaining the abstract and idealized aspects of IST and the intentional stance, I believe that I have filled the gap in Dennett's theory while remaining broadly consistent with his theory.

Bibliography

Akins, K. (1996). Lost the plot? reconstructing Dennett's multiple drafts theory of consciousness. *Mind and language*. 11 (1):1-43.

Akins, K & Dennett, D. (2008). Multiple drafts model. *Scholarpedia*. 3(4):4321

Anscombe, E. (1963). *Intention*. Oxford: Basil Blackwell

Brook, A. (1995). Explanation in the hermeneutic science. *International journal of psycho-analysis*. 76.3:519

Brook, A. (2000). Judgments and drafts eight years later. In Ross, D. & Brook, A. (Ed.). *Dennett's philosophy: a comprehensive assessment*. (pp.219-257). Cambridge, Massachusetts: MIT Press.

Chalmers, D. J. (2014). Intuitions in philosophy: a minimal defense. *Philosophical Studies*. 171 (3):535-544.

Dennett, D. (1978). Intentional systems. In *Brainstorms*. (pp.3-22). Cambridge, Massachusetts: MIT Press.

---------- (1987a). True believers. In *The intentional stance*. (pp.13-35). Cambridge, Massachusetts: MIT Press.

---------- (1987b). Three kinds of intentional psychology. In *The intentional stance*. (pp.43-69). Cambridge, Massachusetts: MIT Press.

---------- (1987c). Intentional systems and cognitive ethnology. In *The intentional stance*. (pp.237-269). Cambridge, Massachusetts: MIT Press.

---------- (1991). *Consciousness explained*. Boston: Little, Brown and.

---------- (1998). *Brainchildren*. Cambridge, Massachusetts: MIT Press.

---------- (1998a). Real patterns. In *Brainchildren*. (pp.94-120). Cambridge, Massachusetts: MIT Press.

---------- (1998b). Do animals have beliefs? In *Brainchildren*. (pp.323-333). Cambridge, Massachusetts: MIT Press.

---------- (2009). Intentional systems theory. In Beckermann, A. *et al.* (Ed.). *The Oxford handbook of philosophy of mind*. Oxford: Oxford University Press

Gregory, R. (2004). Entry on introspection. In *Oxford companion to the mind*. (p.485). Oxford: Oxford University Press.

Hirstein, W. (2005). *Brain fiction: self-deception and the riddle of confabulation*. Cambridge, MA: MIT Press.

Nisbett, R. & T. Wilson (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review* 84(3): 231-259.

Pinel, J. (2011). *Biopsychology*. Boston: Pearson.

Searle, J. (2004). *Mind: a brief introduction*. Oxford: Oxford University Press.