Being Interpreted:

Third-Person Perspectives on the Self

by

Sean Patrick Finn

A thesis submitted to the Department of Philosophy

in conformity with the requirements for

the degree of Master of Arts

Queen's University

Kingston Ontario, Canada

December, 1999

Abstract

Daniel Dennett attempts to provide an adequate explanation of action and agency using the strict third-person methodology of science. He describes our ability to interpret behaviour as a theoretical predictive strategy that involves the postulation of abstract intentional states. Dennett argues that this method can also be applied to understand the phenomenological character of our mental lives. This forces Dennett to describe agency by negotiating the seemingly contradictory fact that agents are physical systems in a deterministic universe that understand themselves using conceptions of value, freedom, autonomy, and responsibility.

I contrast this methodological approach to agency with Charles Taylor's hermeneutic approach to the self. Taylor argues that agency must be construed as a derivative of ethical reflection. This involves understanding the self as dialogical in nature, entailing that the self consists as a process of internalizing and interpreting the perspectives of others. By noting the relative importance of the first-person perspective, Taylor's methodology is at odds with Dennett's approach.

In the final analysis, Dennett's account of agency undermines his entire methodology. The naturalistic presuppositions underlying a strict third-person perspective fail to give adequate recognition to the intimate relationship we have with ourselves as agents. Nowhere in Dennett's account is there room for what matters to us. I conclude that Taylor is right to maintain that such ethical considerations are what make us agents in the first place.

Acknowledgements

It would have been impossible to complete this thesis without the benevolent supervision of David Bakhurst. I wish to thank David for his patience, constructive criticism, and insight, while motivating me to complete this project. I would also like to thank my second reader, Deborah Knight, for her original conception of this thesis, and for performing her role as copy editor with vigour and unrelenting exactness. My internal examiner Lorne Maclachlan, and external examiner Sandra den Otter, also deserve recognition for their comments, questions, and critiques during the thesis examination.

I would also like to acknowledge my family and friends for their continuous love and support. Specifically, I would like to thank my parents and my brother, whose emotional and financial generosity has allowed me to continue my education with a smile. Most importantly, I would like to thank Ivana Dragicevic. Her presence in my life is a continuous source of inspiration, and without her love, this project would never have been completed. Volim te, Ivana.

Table of Contents

Introduction

This thesis examines a general view in the philosophy of mind and action that is best-termed "interpretationalism." Interpretationalism is the view that the attribution of mental states to human beings, and hence the explanation of actions, essentially involves processes of interpretation in which the contents of those mental states and the character of those actions are determined. According to interpretationalists, this is equally true of the attribution of mental states to ourselves as well as to other people; that is to say, it applies to the way we make sense of our own minds as well as the minds of others. A strong interpretationalist position may even argue that these interpretive processes somehow constitute the very existence of mental states.

I will examine the work of Daniel C. Dennett and Charles Taylor, both of whom approach the question of how we account for human mind and action from an interpretationalist stance. Dennett and Taylor reject a variety of opposing views concerning psychology and agency, most obviously behaviourism and mentalism, but also psycho-physical reductionism: the idea that a complete explanation of human mentality can be offered either by appeal to physical laws or neurophysiology. Both share a commitment to our everyday forms of psychological explanation – or "folk psychology" as the explanations are sometimes collectively called – which explain action in terms of "intentional" states, i.e. the beliefs and desires of agents.

What brings Dennett and Taylor together as interpretationalists is their commitment to a third-person perspective towards action. Their point is not the obvious one that because the contents of other minds are unavailable for direct observation, we are obliged to take an external perspective towards other people's behaviour. Rather, they argue for

the view that full human agency, and the reliable description and explanation of human actions, involves incorporating a third-person perspective into any first-person account.

The positions of Dennett and Taylor contrast significantly in the attitude each takes towards the third-person perspective. Dennett represents our everyday intentionalistic psychology from the third-person perspective of empirical science, as a device to be employed in the explanation of behaviour. On this view, the method exercised in explanations of human agency and psychology attempts to make sense of us conceived essentially as physical systems in a physical reality. For Taylor, agency is properly understood through the process of hermeneutic and, ultimately, ethical reflection. According to Taylor, the third-person perspective we must take towards ourselves in our ethical deliberations and evaluations necessarily entails internalizing the first-person perspective of others. Where Dennett and Taylor differ most dramatically therefore, concerns (i) the relative importance of the first-person perspective *vis-à-vis* the third-person perspective when describing intentional action, and (ii) the centrality or marginality of moral concerns as fundamental (or not) to such descriptions.

In this thesis, I argue that Dennett's methodology undercuts the plausibility of his own account of what an adequate description or explanation of human agency should involve. By placing moral concerns to the side, Dennett leaves out what is most important to understanding human agents: *what matters to them as agents*. I will argue, following Taylor, that moral personhood is not just one dimension of human agency among others, but *central* to what makes a human being an *agent* at all. It will be shown that self-understanding and other-understanding are inter-related; one cannot, as Dennett does, posit the understanding of *others* first, with *self*-understanding added on afterwards.

The result of such a methodology distorts the self to such an extent that human agents can no longer recognize themselves in Dennett's account.

Chapter 1

From the Intentional Stance to Heterophenomenology:
Dennett on How to Interpret (People)

1.1 The Intentional Stance

Dennett's understanding of what it is to be a human agent derives from his theory of action interpretation. This section explores Dennett's systematic view of what is involved in an adequate description and explanation of actions. Dennett argues that we can apply the same general method of interpretation to explain all actions in terms of the language and conceptual framework of folk psychology.[1] I will examine Dennett's arguments for this claim, and outline the main presuppositions behind it.

Dennett's commitment to common-sense psychology follows from his belief that the many things that need to be said about the mind cannot be said in the restricted languages of neuroanatomy, neurophysiology, or behaviouristic psychology.[2] It is a philosophical commonplace that our conception of ourselves as "minded beings" – as "responsible, free, and rational agents" – seems to be in tension with our conception of what it is to be a complex part of the physical world of science.[3] Yet Dennett argues that this incompatibility is only apparent; the mentalistic predicates used when speaking about the mind are compatible with the theoretical strategies of legitimate sciences, for example, artificial intelligence research and cognitive psychology.[4] According to Dennett, understanding human action and agency using the language of folk psychology

---

[1] Broadly speaking, folk psychology is a conceptual framework used by people in everyday circumstances to predict, explain, and understand the behaviour and mental states of themselves and others in terms of the states philosophers call "propositional attitudes" (e.g. belief, desire, intention, etc.).
[2] Dennett, "Setting Off on the Right Foot," *The Intentional Stance* (Cambridge: The MIT Press, 1987), 1.
[3] Dennett, *Brainstorms* (Montgomery, VT: Bradford Books, 1978), x.
[4] *ibid.*, xvii.

does not need to be regarded simply as an imprecise, common-sense, rule-of-thumb approach. On the contrary, folk psychology is a legitimate theory for the proper explanation and description of human behaviour. This approach, which Dennett terms the *intentional strategy*, is a theoretical, predictive strategy, by which the behaviour of an *intentional system* is explained by describing that system as possessing certain *intentional states*,[5] i.e. beliefs, desires, and other propositional attitudes.

The intentional strategy is best understood when contrasted with the *design* and *physical* stances that can also be employed to describe and explain the behaviour of a system. Take, for example, Dennett's favorite illustration of the chess-playing computer. One way the behaviour of a chess program can be accurately predicted is simply by knowing how the computer's program has been designed to function. This is a prediction from the *design stance*. If the computer functions without mechanical failure, then accurate predictions of the computer's responses can be made from the design stance if we understand that the computer has been designed to play chess – that it has been programmed to employ various openings, and an array of attack and defense strategies. To take another illustration, consider the behaviour of the human heart. We explain the behaviour of the heart by saying that its *function* is to pump blood and it has been "designed" by evolution to do so.[6] The essential feature of the design stance then, is that behavioral predictions do not have to refer to the specific physical mechanisms that, in these cases, constitute computers and hearts.

---

[5] An intentional state is a state of mind that is *directed at* or *about* something.

[6] The assignment of "function" to the heart involves a criterion of assessment that is internal to the phenomenon under a description from the design stance, but not from a description that regards the heart as simply a physical object. With the assignment of function, there is an assignment of value. See John R. Searle, *The Construction of Social Reality* (New York: The Free Press, 1995), 13-23.

In contrast, to make predictions based upon the physical composition of a system is to be positioned in the *physical stance*. It is theoretically possible to predict the response of a chess computer, or the behaviour of a human heart, by appeal to purely physical laws. If we knew enough about the physical structure of the system at issue, we could determine the state it was about to enter simply by appeal to physical laws. An essential feature of the physical stance then, is that it describes behaviour as essentially a causal process. Of course, performing the necessary calculations would often require an exorbitant amount of time.

According to Dennett, a third strategy is possible when a system's behaviour is too complex to be dealt with effectively from the other stances. This is the *intentional stance*. To predict the behavior of a chess-playing computer from the intentional stance, we assume the rationality of program design and the absence of malfunction in order to make a prediction based upon what we believe the computer *ought* to do if it is to make a reasonable response. In other words, we can treat the computer like an intelligent human opponent. According to Dennett, if a system is consistently predictable via the intentional strategy, it can properly be described as having *beliefs* and *desires*; that is to say, it can be described as an *intentional system*.[7] For Dennett, the advantage of the intentional stance lies with the fact that we can treat any system predictable by this strategy as intentional, without making reference to the system's composition, origin, or status in the moral community; nor should we concern ourselves with notions of consciousness, self-determination, etc.[8] The intentional stance can be used for suitable

---

[7] Dennett, "True Believers," *The Intentional Stance*, 15.
[8] Dennett, *Brainstorms*, 7.

"agents" precisely when we are unable to deploy the physical and design stances. It is a reliable technique for explaining the behavior of what are (ultimately) physical systems.

Dennett describes the process by which we formulate our general theories of interpretation as follows:

> ...first you decide to treat the object whose behaviour is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs.[9]

Let us elaborate upon this. In contrast with chess-playing computers designed by human intelligence, consider organic systems "designed" by the processes of evolution. What is entailed in figuring out what beliefs a system ought to have, and what does it mean to use its place and purpose in the world as a guide? According to Dennett, to deduce what beliefs a system ought to have, we "attribute as beliefs all the truths relevant to the system's interests (or desires) that the system's experience to date has made available."[10] Dennett also tells us that a system has mostly true beliefs, because simply being exposed to certain states of affairs over a period of time is usually the sufficient condition for coming to know those states of affairs.[11]

The beliefs we ought to attribute to a system are, Dennett argues, closely related to the desires a system ought to have. In fact, deducing the desires a system ought to have is central to the intentional strategy, because only from these can we shape the list of beliefs

---

[9] Dennett, "True Believers," 17.
[10] Dennett, "True Believers," 18.
[11] *ibid.* This is similar to conclusions reached by Donald Davidson from a different route. According to Davidson, a person's beliefs must be "largely true" because we could not interpret someone unless this were so. Unless we assume that the person we are interpreting has largely true beliefs we do not know how to limit the hypotheses about what they might mean. This precondition of interpretation, i.e. of the possibility of language, is called the Principle of Charity. See Donald Davidson, "A Coherence Theory of

that ought to accompany them. The fundamental rule, according to Dennett, is to attribute the most basic desires to subjects – "survival, absence of pain, food, comfort, procreation, entertainment."[12] Predictions are then based upon what the system will do to further these "goals" in light of the accompanying beliefs.

When a prediction is confirmed, we are better able to determine what more specific beliefs and desires the system has. These concepts allow Dennett to explain the observed behavior of a system in a way that the physical and design stances cannot. The intentional strategy has the benefit of being a quick and easy way of processing information. From an evolutionary perspective, the physical and design stances are practically inaccessible. For example, in the time it takes to finish our prediction of what a hungry tiger will do from either the physical or design stance, we will be eaten.

One of Dennett's key background assumptions is the concept of behavioural patterns. Among the reasons the intentional stance is so successful, Dennett believes, is because any confirmed prediction has relied upon existing patterns of behavior that are genuinely "real." Complex systems form complex patterns, and in this sense the patterns are objective since there is something to be detected. But these patterns are not entirely independent of us, since they require beings like us to recognize them in the first place. Thus, the reason the intentional stance is so successful as a predictive strategy is because human beings are constituted in such a way as to easily recognize intentional patterns, and it is the recognition of these patterns that makes the prediction of intentional behaviour possible at all. Dennett argues that the intentional stance can only work

---

Truth and Knowledge," reprinted in E. Lepore (ed.) *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, (Oxford: Basil Blackwell, 1986), 307-319.

[12] Dennett, "True Believers," 20.

because such patterns are real, and that someone who makes predictions only from the physical or design stances would miss the patterns discernable from the intentional stance.[13] This is not because the physical and design stances are somehow less effective (in theory) in predicting the behaviour of a given system. Rather, it is because the intentional strategy has a unique ability to penetrate the world in a particular way. Thus, it is because intentional patterns are "composed partly from our own 'subjective' reactions to what is out there"[14] that folk psychological concepts are the only tools that can be employed properly to explain the observed, *intentional* behaviour of a system.

For example, when describing the world from the physical stance, certain patterns become obvious, but when describing the world in any other way, the same patterns become invisible, and new ones come to light.[15] From the physical stance, we describe the behavior of a system making reference only to causal facts. But when the system is a human being, and the behavior being described is an action, the concepts of cause and effect become limited, if not useless,[16] and we must resort to an alternative strategy to explain this behavior properly.

Understanding the idea of a real pattern requires that we also recognize the incompleteness and imperfection in any intentional pattern. Two important points about patterns follow from this. First, it is possible correctly to predict the behaviour of a system for the wrong reasons. If we explain the behaviour of a system using the concepts

---

[13] *ibid.*, 25.
[14] Dennett, "Real Patterns, Deeper Facts, and Empty Questions," *The Intentional Stance*, 39.
[15] *ibid.*, 39n.
[16] Behaviour can only be considered an action if the purpose of or intention behind the action is assumed, otherwise there would not be a difference between actions and the mere movement of physical bodies through space. Furthermore, intentionality cannot be reduced to physical laws, for this would require that we can state necessary and sufficient conditions for action in intentional terms, but no purpose, goal, or intention has the power to compel or direct our behaviour in such a way. A causal description of action only makes sense when actions are teleologically described, but physical laws are not teleological.

of folk psychology, it is always possible to tell another story using different intentional states to accommodate the same behaviour. Since "[t]he objective presence of one pattern (with whatever imperfections) does not rule out the objective presence of another pattern (with whatever imperfections)"[17] our correct prediction of a system's behaviour may in fact rest on a misunderstanding of the pattern.[18] Second, we may predict the behaviour of a system incorrectly even though we had the correct rationale. The possibility of a system failure for whatever reason (i.e. malfunction, design imperfection, irrationality, etc.) will lead inevitably "to circumstances beyond the power of the intentional strategy to describe."[19] It follows from these two points that any explanation of action from the intentional stance will always be an *interpretation* of a system's behaviour.

An intentional system is anything that can be consistently predicted by the intentional strategy, and anything predicted as such is said to possess certain beliefs and desires. Dennett claims that as a theory of action interpretation, the intentional stance and its use of folk psychology is as legitimate as any other theoretical strategy for the explanation of behaviour. This analysis of action interpretation obviously brings forth questions of what it is like to be a being that engages in this kind of interpretive strategy. Specifically, what kind of beings must we be in order to produce, recognize, and easily interpret intentional patterns? Before we turn towards Dennett's answer to this, we need to explicate an important background assumption that has a direct bearing on such questions.

---

[17] Dennett, "True Believers," 29.

[18] Such a possibility is considered "practically negligible" by Dennett, since he believes that the more complex the behaviour, the less chance there is of a dual, systematically unrelated interpretation. See *ibid.*, 29n.

1.2 The Assumption of Rationality

When using the intentional strategy, we are to treat the object whose behaviour is to be predicted as a rational agent. What exactly does this require? It requires the assumption of rationality. According to Dennett, assuming rationality entails that we take it for granted that agents believe "all the implications of their beliefs and believe no contradictory pairs of beliefs."[20] It seems however, that this takes us away from the question of what is involved in an *accurate* description or explanation of human action, since it is obvious that no human being is ideally rational.[21] If this is the case, then the intentional stance may seem inadequate because it is premised on a false assumption. Stephen Stich has raised this objection.[22] If human agents are prone to make mistakes in their reasoning, then the intentional stance and its assumptions of ideal rationality will fail to incorporate important aspects of what it is like to be a human agent, for example, human error, idiosyncrasies, and other cognitive shortcomings, as well as the potential for cognitive growth.[23]

While Dennett does state that we should assume systems to be ideally rational, he also admits that "[n]o one is perfectly rational, perfectly unforgetful, all-observant, or invulnerable to fatigue, malfunction, or design imperfection."[24] Although this statement seems to contradict the one quoted above, what Dennett is trying to argue is that the assumption of rationality is not a description of what a system is *actually* like. The

---

[19] *ibid.*, 28.
[20] *ibid.*, 21.
[21] For examples and reasons why not, see Stephen P. Stich, "Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality," *Naturalizing Epistemology*, Hilary Kornblith, ed. (Cambridge: The MIT Press, 1985), 249-266.
[22] Stich, "Dennett on Intentional Systems," *Philosophical Topics* 12 no.1 (Spring 1981).
[23] *ibid.*, 48.
[24] Dennett, "True Believers," 28.

assumption of ideal rationality in Dennett's theory should be thought of as analogous to the assumption of frictionless surfaces used in the study of mechanics. Consider Dennett's definition of rationality:

> I want to use "rational" as a general-purpose term of cognitive approval – which requires maintaining only conditional and revisable allegiances between rationality, so considered, and the proposed (or even universally acclaimed) methods of getting ahead, cognitively, in the world.[25]

The conception of rationality found in Dennett's intentional strategy needs to be adapted to real world circumstances. This is because the circumstances which dictate "how one gets ahead cognitively in the world" can vary from system to system. Dennett's definition of rationality is a variation of the fundamental rule of interpretation stated earlier: To attribute the desires and beliefs a system ought to have, given its place in the world and its purpose. Not only does a system have certain basic wants (for food, warmth, health, pleasure, sex, companionship, etc.) it also has the capacity to acquire new wants, broadly compatible with these basic wants. Human beings are creatures that can even consciously develop desires (e.g. for chastity) that conflict with their basic wants. Since these aims may be achieved in a number of ways by a variety of systems in various circumstances, then to predict behavior accurately one must begin with ideal rationality, that is, what the system *ought* to believe and desire, and then revise downward as circumstances dictate.[26]

By revising downward from our initial assumptions of ideal rationality, Dennett believes we can accommodate our all-too-human tendency to be less than wholly rational. To revise one's theory from within the intentional stance, one must either adjust one's theory concerning what the system believes, or re-evaluate what the system's goals

---

[25] Dennett, "Making Sense of Ourselves," *The Intentional Stance*, 97.

and desires are. The test case is mistakes. According to Stich, we do not usually "revise downward" when we discover that a person has been less than ideally rational, for example, when they have made a mistake. We simply assume that the person has *made a mistake*, that they made a blunder in their reasoning.[27] How does this objection impact the intentional stance?

Suppose, using an example of Stich's, that I purchase a glass of lemonade from a boy who is selling it for twelve cents a glass. Also suppose that after handing him a quarter, the boy gives me a glass of lemonade, and a dime and a penny change. Now, when we point out the error to him, Dennett argues that we can *predict* a number of reactions:

> That he will exhibit surprise, blush, smite his forehead, apologize, and give me two cents. Why do we expect him to exhibit surprise? Because we attribute to him the belief that he's given me the right change - he'll be surprised to learn that he hasn't. Why do we expect him to blush? Because we attribute to him the desire not to cheat (or be seen to cheat) his customers. Why do we expect him to smite his forehead or give some other acknowledgement of his lapse? Because we attribute to him not only the belief that 25 - 12 = 13, but also the belief that that's obvious, and the belief that no one his age should make any mistakes about it.[28]

But if we now want to describe *the error* using the language of folk psychology, then according to the intentional stance, we must revise downward from our original assumption of ideal rationality. Here we run into difficulty. Which belief is to be altered? We know that the boy believes that he has given me the right change, but does he not believe that I gave him a quarter?; that his lemonade costs twelve cents?; that a

---

[26] Dennett, "True Believers," 21.
[27] Stich, "Dennett on Intentional Systems," 50.
[28] Dennett, "Making Sense of Ourselves," 84.

quarter is 25 cents, a dime 10 cents, a penny 1 cent?; that he gave me a dime and a penny change?; that 25 - 12 = 13?; that 10 + 1 = 11?; that 11 ≠ 13?[29] It is difficult to understand how the boy had the *false* belief that he had given me the right change, given all of these other *true* beliefs. It would not have been possible to attribute the false belief to the boy in our initial predictions either, since the notion of what a *mistake* is would then be lost. That is, if we had a complete and accurate description of a rational thinker with a single false belief, no error could be described as having been made in the first place.[30] Are we to conclude then, as Stich does, that this action of giving the wrong change elicits no description from the intentional stance?

By its very nature, the intentional stance will have difficulty explaining and describing irrational behaviour. For Dennett however, this is not a failing of the intentional stance but a result stemming from the meaning of our folk psychological terms. For Dennett, "belief" and "desire" are terms used to describe a system whose behaviour can best be explained by the possession of certain information (or misinformation[31]). However, it is impossible to give a rationalizing explanation of the lemonade seller's error using folk psychological terms. Attempting to discover what his *belief-set* must have been to make the kind of mistake he did is an attempt to render the mistake intelligible under an explanatory scheme of rationality, but with pure errors there is no rationale to be found.

---

[29] *ibid.*, 85.
[30] *ibid.*, 86.
[31] There are mistakes that can be explained by the intentional stance, for example, mistakes that rest on misperception. We can easily imagine a scenario whereby a lemonade seller mistakes a nickel for a quarter, in which case the intentional strategist would proceed to explain the behaviour by incorporating a false belief into the lemonade seller's total belief set.

Dennett therefore concludes that we must descend from the level of beliefs and desires to some other level of theory to explain mistakes, since no account in folk psychological terms will make complete sense. However by switching to the physical or design stance, the description of the agent and its actions are reduced to patterns involving causal mechanisms and processes. We therefore lose our ability to use *intentionality* as an explanation of action. Thus, Dennett believes that we must think of "belief" and "desire" as *sense-making terms*. The intentional strategy is *our* way of "making sense" of a system's behaviour. Since mistakes are actions that do not make sense they are *unreasonable* by definition, and they therefore cannot be properly expressed within an explanatory scheme of rationality.[32]

So Stich is correct in claiming that the intentional stance does fail to provide certain information about what it is like to be a human agent, *especially* facts concerning human error, idiosyncrasies, shortcomings, and the potential for cognitive growth. This is not a concern for Dennett, however, because the intentional strategy is a method to be used for the proper description and explanation of *intentional actions*. A mistake, by definition, cannot be considered an action proper, since any behavior described as an action presupposes the notion of goals, purposes, or intentions, which is precisely what mistakes necessarily lack.

For Dennett, a prediction is made with assumptions concerning what a system ought to believe and desire, and what it rationally ought to do to further its goals. If the prediction fails, adjustments will need to be made in the original assumptions. Through the process of trial and error, one can eventually get the predictions right, and for

---

[32] Dennett, "Making Sense of Ourselves," 87.

Dennett, this says something important about the system itself as it confirms what beliefs and desires the system can be said to possess. In the process, one learns something important about the system in question, whether that system is a spider, a frog, a computer, or a lemonade seller. Through the methodology of the intentional stance, we come to *know* the agent's world as she "sees" it. This for Dennett, it seems, is all that is necessary for understanding a particular system as an agent. Thus, it is Dennett's conclusion that this same process can be used to understand what it is to be a human agent.

1.3 True Believers: Do We Really Have Beliefs and Desires?

Dennett argues that we can adopt the intentional stance toward many different intentional systems: spiders, frogs, computers, thermostats, human beings. While we can *describe* spiders, frogs, computers, etc., as having beliefs and desires, surely it is only human beings who are "true believers." We believe that only human beings *really* have beliefs, and that talk of propositional attitudes with regard to non-human systems is just anthropomorphism. What then, is the status of propositional attitudes described from the perspective of the intentional stance?

When we describe human beings as *having* beliefs and desires, we seem to mean it in a non-metaphorical way. But contrary to the view that beliefs and desires are *things* to be possessed, Dennett argues that propositional attitude descriptions have an ontological equivalence to the theoretical posits of science. Intentional predicates (such as beliefs and desires) are explanatory devices that are projected onto systems in order to describe and explain their behaviour. According to Dennett: "Folk psychology is thus *instrumentalistic* in a way the most ardent realist should permit: people really do have

beliefs and desires, on my version of folk psychology, just the way they really have centers of gravity and the earth has an Equator."[33] Intentional attitudes provide important information about cognitive systems, but they do not pick out any *internal states* of those systems, just as "centers of gravity" and "the Equator" do not pick out any *thing* in the world. Therefore, the status of propositional attitudes when applied to human beings is the same for any intentional system. On Dennett's view, "[human] belief and desire are like froggy belief and desire *all the way up*."[34] That is to say, when we use "belief talk" and "desire talk" to describe ourselves and each other as human beings, we are saying no more than when we describe spiders, frogs, and computers as having such states.

If propositional attitudes are theoretical posits, then it may be objected that the intentional stance is too broad in its application to be of any value. Surely the intentional stance cannot be applied to just *any* system. As it has been described thus far, however, it does seem that Dennett's strategy fails to exclude anything whatsoever. If Dennett's methodology can be applied to objects such as rocks or lecterns, then does this not constitute a *reductio ad absurdum*? For example, a lectern could be construed as believing itself to be in the center of the room, and desiring to remain there, and thus, from the intentional stance, it seems we could then accurately predict that the lectern would not move.[35] But when (of course) the lectern does not move, must we now grant the lectern certain beliefs and desires? If the intentional stance is to have any merit as a legitimate predictive strategy, Dennett must provide a way of rejecting certain kinds of systems as being properly *intentional* systems. Otherwise, the meaning of "belief,"

---

[33] Dennett, "Three Kinds of Intentional Psychology," *The Intentional Stance*, 52-53.
[34] Dennett, "When Frogs (and Others) Make Mistakes," *The Intentional Stance*, 112.
[35] This example is taken from Dennett, "True Believers," 23.

"desire," and "intention" will be diffused to a point where they could serve no instrumental purpose.

To determine what *really* has beliefs and what does not, Dennett draws a comparison between thermostats and human beings. According to Dennett, the reason why systems like human beings can properly have beliefs and desires attributed to them, while a thermostat cannot, is because there is not a strong enough link between the thermostat and the world.[36] Regardless of what environment or circumstance the thermostat is placed in, it will function in exactly the same way. For Dennett, the attribution of beliefs and desires to a given system must rely heavily upon the way that system behaves in a variety of circumstances. If a system's performance fails to change under varying circumstances, Dennett believes we should be hesitant in our attribution of propositional attitudes. In order for a system to be described as having intentions, it must have an *awareness* of its environment such that we can describe it as having *formulated* different desires from within differing states of affairs. Along with predictability, intentional systems must also have certain "modes of attachment" unavailable to functionally impoverished systems like thermostats.[37]

A thermostat would have to enrich its modes of attachment to the world to warrant an intentional description and according to Dennett, this would require "vast complications of its inner structure."[38] If we could imagine a thermostat, as Dennett suggests, that was able to *choose* its own boiler fuel, *purchase* it at an economically desirable rate, and *check* the weather stripping at the same time, then the actions of

---

[36] *ibid.*, 30.
[37] *ibid.*
[38] *ibid.*

choosing, purchasing, and *checking* are patterns of behaviour that can be recognized as warranting an intentional description. With regards to the new, refurbished thermostat, Dennett says:

> This adds another dimension of internal complexity; it gives individual belief-like states *more to do*, in effect, by providing more and different occasions for their derivation or deduction from other states, and by providing more and different occasions for them to serve as premises for further reasoning.[39]

The old thermostat could be placed in any environment and it would do exactly the same thing it had always done. Its function is singular, and it could not be expected to function differently in radically diverse environments. Systems like the refurbished thermostat – or a human being – are complex mechanisms that are able to form elaborate patterns because of their internal complexity. "If you change [a system's] environment, it will *notice*, in effect, and make a change in its internal state in its response."[40] We know a system *really* has beliefs and desires then, because this is the only way a system would be able to react to its environment in such a way that it could elicit a behavior capable of being interpreted as "getting ahead in the world." The system must be, in some sense, *aware* of its surroundings for it to be described as *acting* in a certain way. The action can then be *explained* using folk psychological terms.

Human beings are "true believers" because they have a complicated inner structure that is properly engaged with the world. This internal complexity allows us to respond to our environment in such way as to elicit a complicated pattern of behaviour that can be recognized and described by other competent observers as intentional. But human beings are therefore just one system among many that can warrant a description of *having* beliefs

---

[39] *ibid.*
[40] *ibid.*, 31.

and desires. While human beings are able to express very complicated beliefs and desires – due to their ability to use language – the intentional predicates ascribed refer to no deeper facts than those attributed to spiders, frogs, and computers. All intentional systems are "true believers" according to Dennett.

1.4 Dennett's Third-Person Perspective

With his insistence upon the description of behaviour from the third-person perspective of science, we can classify Dennett as a kind of "post-behaviourist." "Inner" mental states are not excluded from his psychology, as was the case with traditional behaviourism, but are kept in play as constructs the intentional theorist posits to explain the observed behaviour. If mental states are theoretical posits directly related to the physical constitution and behaviour of a system, then it may be argued that Dennett is committed, like the behaviourists before him, to denying, or at least downplaying, the phenomenological character of our mental lives. To say that beliefs and desires are the theoretical posits of a quasi-scientific method is one thing, but is the intentional strategy supposed to apply to our sensations, moods, and emotions as well? Where exactly does the intentional stance leave the issue of phenomenology?

As we have seen, Dennett employs a strict third-person perspective, and he continues to insist upon the third-person perspective when he turns to consider phenomenology. This is because, for Dennett, a serious examination of phenomenology requires the adoption of agreed-upon methods of description and analysis, and naturally, Dennett finds such precision in the third-person perspective of science. According to

Dennett, the methodology of science allows himself to be neutral[41] about two important debates: i) about the subjective and objective approaches to phenomenology, and ii) about the physical or non-physical reality of phenomenological items. I will examine each point in turn.

First, Dennett rejects the subjective, first-person approach to the mental in favor of an objective, third-person perspective of science because what the first-person perspective lacks is a well-established and agreed-upon method of describing and analyzing the phenomenological aspects of the mind. Ever since Descartes wrote his *Meditations* in the first-person singular, it has been commonly accepted that the sense of immediacy a person has with respect to their mind entails privileged access to their own mental states, and therefore special authority over their content. Not only that, what is "observed" when one introspects is also assumed to be universal – every person seems to know and agree upon *what it is like* to be a "minded human being." However, while Dennett does agree that a person has authority over what the contents of their mind might *be* (phenomenologically speaking), he disagrees as to what authority their subjective position *represents* in actuality. And while we all have the ability to introspect and discuss what we "see," Dennett believes that "controversy and contradiction bedevil the claims made under these conditions of polite mutual agreement."[42]

---

[41] Although it is Dennett's term, it does seem odd to describe the methodology of science as "neutral," since it dismisses the first-person perspective out of hand. This is explained by Bo Dahlbom: "Dennett's behaviourism and functionalism are related to verificationism, the idea that where there can be no evidence to decide an issue, there is no issue. The notion of evidence here is borrowed from science: whatever natural scientists in their practice will count as evidence, Dennett will accept." Thus, Dennett's methodology is "neutral" only from the perspective of science, which is the only relevant perspective for Dennett. See Bo Dahlbom, ed., *Dennett and His Critics*, (Cambridge: Blackwell Publishers, 1993), 5.

[42] Daniel C. Dennett, *Consciousness Explained* (Toronto: Little, Brown & Company (Canada) Limited, 1991), 67.

Dennett's main reason for dismissing the first-person perspective is his contention that it "is a treacherous incubator of errors."[43] Dennett cites numerous scientific and psychological experiments that show that the assumptions of an authoritative first-person perspective, and the accuracy of its representations, are misguided. For instance, before the rise of scientific inquiry, it was not known that the human visual field has blind spots, that we "hear" the sound of bass through our feet, that we taste with the help of our noses, and that we have an enormous deficiency in our peripheral vision.[44] Likewise, in experiments displaying what is known as the colour phi phenomenon,[45] for example, our *perception* of what is occurring (what it *seems like* to us) is significantly different from what is *actually* occurring. According to Dennett, a pre-theoretical account of phenomenology that focuses only upon experiences, and not upon the causes and effects of those experiences, is significantly lacking as a complete explanation of phenomenology. Not only does a focus on causes and effects allow for external criteria of assessment to be established, it also allows for a careful, inter-subjective collection of information.

This is *not* to champion the third-person perspective as an objective stance claiming to know and tell us *what it is like* to be a "minded human being." According to Dennett, one has to wait and see what a scientific theory says before either accepting or rejecting it.[46] However, the scientific methodology is the only approach to the mental that allows

---

[43] *ibid.*, 70.
[44] *ibid.*, 68.
[45] Originally conceived by Max Wertheimer (1912), but adopted by Paul Kolers and Michael von Grunau (1976). Two distinct coloured spots are lit for 150msec each with a 50msec interval. The phenomenological account given by the subject is that the first spot *seems* to begin moving and then changes colour abruptly in the middle of its illusory passage towards the location of the second spot. For this and other experiments, see Dennett, *Consciousness Explained*, 114.
[46] *ibid.*, 71.

phenomenological data to be collected and described without a bias towards what it might actually represent in the "real world." Dennett believes that observing the behavior of subjects, recording what they say about their experiences, and then organizing and analyzing the data using the principles of science, is the only way the "objects" of phenomenology can be carefully catalogued without prejudging the issues of the debate.

Dennett believes the third-person perspective also allows him to be neutral with respect to the physical and non-physical reality of phenomenological items. So ontological and metaphysical commitments concerning various mental events are set aside. Although Dennett is proposing a theory of mind from the third-person, materialistic perspective of contemporary science, folk psychology is being used *abstractly*, "in that the beliefs and desires it attributes are not – or need not be – presumed to be intervening distinguishable states of an internal behaviour-causing system."[47] Just as the concept of belief has the same explanatory role as centers of gravity ("the calculations that yield the predictions are more like the calculations one performs with a parallelogram of forces than like the calculations one performs with a blueprint of internal levers and cogs"[48]), so phenomenological items are products of interpretation, and their "reality" rests only in their role in the prediction and explanation of behavior. Using the scientific method, one tests various hypotheses through prediction and experimentation in order to develop a working theory of the mind.

1.5 Heterophenomenology

How is the third-person perspective of science to understand the phenomenological

---

[47] Dennett, "Three Kinds of Intentional Psychology," 52.
[48] *ibid.*

character of our mental lives? Following William James, Dennett believes that to do psychology one needs to determine the relationship between "the *semantic* properties of the internal events and structures under examination" (the subject's "mental representations") and the subject's environment.[49] The former component seems to indicate that there must be a necessary contribution from the perspective of the subject for any psychological explanation of that subject's actions or experiences. In this regard, Dennett argues that an intentional theorist can posit a "notional world" based upon the behavior (including verbal behaviour) and physical constitution of a subject, which can then be used as a *model* for the subject's internal representations. Dennett explains:

> A notional world should be viewed as a sort of *fictional* world devised by a theorist, a third-party observer, in order to characterize the narrow-psychological states of a subject. A notional world can be supposed to be full of notional objects, and the scene of notional events – all the objects and events the subject believes in, you might say.[50]

Eventually, the notional world of a subject is the world that a subject "believes" in, which we posit, from a third-person perspective in light of the subject's behaviour. According to Dennett,

> [t]he notional world we describe by extrapolation from current state is thus not exactly the world we take to have created that state, even if we know that actual world, but rather the apparent world of the creature, the world apparent *to* the creature as manifested in the creature's current total dispositional state.[51]

Some examples should help to clarify this idea. The psychology of a woman suffering from *Anorexia nervosa*, for example, can be explained from the third-person perspective by postulating a notional world that we interpret the woman's behaviour to

---

[49] Dennett, "Beyond Belief," 154.
[50] *ibid.*, 153.
[51] *ibid.*, 157.

involve. We know that the woman *perceives* herself to be overweight – even though in *actuality* she is not – and we also know that her actions will be based upon this phenomenological experience. We can then predict and explain the woman's behaviour by attributing to her a belief in our constructed notional world. That is, the woman's desire not to eat is explained by her belief in a notional world, the content of which contains, from the third-person perspective, the false perception that she is overweight, and that not eating is the means to satisfy her desire to be thin. This process by which an intentional theorist describes the phenomenology of a subject based upon the behavior observed from a third-person perspective, Dennett calls "heterophenomenology."[52]

Here is another example, this one from Dennett, involving the placement of stereo speakers: Due to the balance of the volume and their respective outputs, a listener will *project* the resulting sound to a point between the two speakers.[53] It is obvious that there are no *literal* projections of audio properties. The point between the two speakers is just a part of the listener's phenomenology. *It just seems to the agent* that the sound is coming from that point. This is what the system has projected into its "phenomenal space,"[54] and is what is described by the heterophenomenologist as an element in the subject's notional world. Thus, the behavior of the agent in this case constitutes a space that enables a third person properly to interpret the actions within it. This includes making predictions about how the agent will behave in light of this space, predictions that will explain the exhibited behavior by making reference to the beliefs and desires the agent has, or seems to have, concerning the world around her.

---

[52] Dennett, *Consciousness Explained*, 72.
[53] The example is from *Consciousness Explained*, 130.
[54] *ibid.*

Thus, according to Dennett's position, proper psychology entails rejecting the unreliable, first-person phenomenological account in favour of an objective, third-person, heterophenomenological account. The first-person perspective is not being ignored on this view, it is simply being subjected to the imagination and criticism of what Dennett believes to be a more trustworthy power. Dennett's analogy between the heterophenomenologist's task of interpreting a subject's behaviour, and the reader's task of interpreting a work of fiction, nicely summarizes the point:

> [O]ur experimenter, the heterophenomenologist, lets the subject's text *constitute* that subject's ***hetero-phenomenological world,*** a world determined by fiat by the text (as interpreted) and indeterminate beyond. This permits the heterophenomenologist to postpone the knotty problems about what the relation might be between that (fictional) world and the real world. This permits theorists to agree in detail about just what a subject's heterophenomenological world *is*, while offering entirely different accounts of how heterophenomenological worlds map onto events in the brain (or the soul, for that matter). The subject's heterophenomenological world will be a stable, intersubjectively confirmable theoretical posit, having the same metaphysical status as, say, Sherlock Holmes's London, or the world according to Garp.[55]

The "subject's text" is the observer-postulated phenomenology of the subject. For Dennett, if a subject's behaviour continuously mirrors the notional environment, then a heterophenomenologist can say that there is an accurate representation of that environment implicit in the organization of the subject.[56] A subject's notional world will overlap, though will never completely agree with the *real* world. The idea of a notional world allows an intentional theorist to posit the many "mistaken identities, chimeras and personal bogeymen, factual errors and distortions" that can be responsible for a subject's actions, especially if that subject is a human being.[57]

---

[55] *ibid.*, 81.
[56] Dennett, "True Believers," 31.
[57] Dennett, "Beyond Belief," 158.

It seems that we have completely explicated Dennett's answer to what is involved in an adequate explanation of action interpretation. It was shown that this entailed a third-person methodological approach to the mental. However, it is still unclear how to describe a human agent on this view. Does this entail applying heterophenomenology to ourselves? If so, how can we do this? And if we can, how does this form of other-understanding affect self-understanding?

Chapter Two

The Evolution of Self and Agency: Mind as Artifact and Narrative

Two strands of thought converge in Dennett's conception of agency: i) the question of the evolution of consciousness, and ii) the issue of determinism and free will. An important question that underlies both is how a physical system like the brain is able to develop a conscious sense of self. That is, how is it possible that a configuration of atoms can become a "minded human being," complete with a sense of self, moral responsibilities, and conceptions of value? The short answer for Dennett is that this is *not* possible, it just *seems* like it is.

Nevertheless, Dennett is not a reductionist when it comes to understanding the concepts of agency and action. A conception of the self that relies heavily on evolutionary theory to account for how phenomena, such as the mind and the self, arise through the evolution of physical structures must be understood in a way that does not reduce them to those structures. However, this conception forces Dennett to accommodate such notions as agency and free will, even though the causally determined[58] universe of evolutionary theory seems to undermine such concepts.

The first two sections of this chapter briefly trace Dennett's understanding of how the self can be incorporated within evolutionary theory. This entails an outline of Dennett's theory of the evolution of mind. The evolutionary story Dennett tells is not intended to be one of fact, but one that is supposed to be plausible, given what we know about the brain, phenomenology, and evolution. The third section then shows why

---

[58] Determinism is the view which holds that all physical events are causally determined by prior events.

Dennett believes it is possible for human selves to be *agents*, in a way that is compatible with our understanding of both the physical world and what it is like to be a conscious human agent.

2.1 The Evolution of Mind and Self

According to Dennett, there were three evolutionary processes required for the development of a conscious mind. The first is the establishment of a boundary between self and world. In order to "avoid the 'bad' things and seek the 'good' things" – necessary for acts of self-preservation – a line between "self" and "world" must be drawn. For Dennett, the "distinction between everything on the inside of a closed boundary and everything in the external world ... is at the heart of all biological processes."[59]

The second process is the establishment of new and better ways of *producing future*.[60] Dennett postulates an evolutionary development of nervous systems from pain receptors, to "hard-wired" automatic response mechanisms (i.e. para-sympathetic nervous systems), to anticipatory components (i.e. sense organs), and eventually, in mammals, to a division of labor in the brain, i.e. left hemisphere/right hemisphere specializations. Eventually, by enhancement through natural selection, which includes ever-increasing brain plasticity, we reach the brain of *Homo sapiens*, the size and shape of which is the same now as it was 10,000 years ago.

Before addressing the third process, it is important to note that according to Dennett, evolution occurs whenever three conditions are met: *variation* in abundance and

---

[59] Dennett, *Consciousness Explained*, 174.
[60] To *produce future* is to "extract anticipations in order to stay one step ahead of disaster." See *Consciousness Explained*, 144.

number of elements; that the possibility of *replication* exists within those elements; and that the number of copies of an element that are created at a given time varies with respect to conditions of the environment and features of the element (*differential "fitness"*).[61] While this is a blueprint for natural selection, it is not necessarily so for genes (DNA molecules) only. On Dennett's story, the development of human consciousness relied, not only upon advances made in brain design, but also upon another set of replicators that fit this scheme: *ideas*.

With this we reach Dennett's third process, the one that makes the development of the mind possible: cultural evolution. Dennett argues that complex ideas formed themselves into distinct memorable units (which he calls *memes*, following Richard Dawkins[62]) through processes of social interaction. These processes of social interaction occurred through particular languages in the form of sayings, pictures, books, as well as buildings, tools, and other artifacts. Basically, any social institution can be considered a *meme-vehicle*. Examples of such complex ideas or "cultural units" are: wheel, vendetta, alphabet, the Odyssey, chess, Impressionism, "Greensleeves," right triangle, etc.[63] The cultural evolution of memes, according to Dennett, follows the laws of natural selection exactly:

> [J]ust as the genes for animals could not come into existence on this planet until the evolution of plants had paved the way ... so the evolution of memes could not get started until the evolution of animals had paved the way by creating a species – *Homo sapiens* – with brains that could provide shelter, and the habits of communication that could provide transmission media, for memes.[64]

---

[61] *ibid.*, 200.
[62] Richard Dawkins, *The Selfish Gene*, (Oxford: Oxford University Press, 1976).
[63] The examples are taken from Dennett, *Consciousness Explained*, 201.
[64] *ibid.*, 202.

Whether or not memes get copied, persist, and multiply, depends on selective forces that act directly on the meme-vehicles that embody them.[65]

The development of language makes specific aspects of cultural evolution possible. According to Dennett, "there was a time in the evolution of language when vocalizations served the function of eliciting and sharing useful information."[66] The "internal dialogue," with which human beings are so well acquainted, arises as a natural side effect of this sharing/communicating with others. Once a brain has built entrance and exit pathways for vehicles of language it then becomes "*parasitized*" by memes, creating an organized structure of complex ideas: *a human mind.* From a certain perspective, in the same way "a scholar is just a library's way of making another library," a human mind is just an idea's way of maintaining and promulgating that idea.[67]

On Dennett's evolutionary story then, the human mind becomes an artifact, created "when memes restructure a human brain in order to make it a better habitat for memes." The greatest difference between individual minds lies not in the physical structure of the brain, but is found in the variation among the memes that make them up. As Dennett notes, "native Chinese minds differ dramatically from native French minds, and literate minds differ from illiterate minds."[68] Thus, a conception of mind cannot even in theory be built by simply examining the structure of the brain, since the mind, and everything that can be said about it, is an artifact constructed by memes:

> It cannot be "memes versus us," because earlier infestations of memes have already played a major role in determining *who or what we are*. The "independent" mind struggling to protect itself from alien and dangerous memes is a myth; there is, in the basement, a persisting tension between

---

[65] *ibid.*, 204.
[66] *ibid.*, 195.
[67] *ibid.*, 200.
[68] *ibid.*, 207.

> the biological imperative of the genes and the imperatives of the memes, but we would be foolish to "side with" our genes ...[69]

Although the brain is subject to physical laws, the incorporation of memes into the brain allows it to produce the *illusion* of mind and the corresponding sense of *agent causation*.[70]

2.2 Mind and Self as Illusion

For Dennett, the brain is a syntactic engine that *mimics* the competence of a semantic engine.[71] When mechanical push comes to shove, a brain will always do what it is caused to do by antecedent physical circumstances. Thus, all a brain can do is "*approximate* responsivity to meanings that we *presuppose* in our everyday mentalistic discourse."[72] We can now see how Dennett believes this to be possible: the brain could be designed by evolutionary processes "to do the right thing (from the point of view of meaning) with high reliability."[73] Thus, there is no such *thing* as *a mind*, but this does not mean that *minds* are not *real*. From Dennett's third-person perspective, the "puzzling legerdemain" of consciousness is the effect of interactions between the brain and memes. This is what allows human beings to have the phenomenological experiences they have. However, the *mind* as an individually existing entity is an illusion. From the third-person perspective, the brain is performing a cognitive illusion that leaves us understanding one another and ourselves *as if* each of us had such an individually existing entity, a "mind."

Dennett's understanding of how the development of a self is possible parallels his

---

[69] *ibid.*
[70] The term "agent causation" contains the idea that conscious human beings are the *source* of actions, not the *effect* of some hidden causal processes. Actions stemming from the human "will" are thought to be elicited from a self-originating source.
[71] Dennett, "Self-Portrait," *Brainchildren*, (Cambridge: Bradford Books/The MIT Press, 1998), 357.
[72] *ibid.*
[73] *ibid.*

story of how the evolution of mind is possible, along with similar conclusions. For Dennett, the human self and all of its facets are simply extensions of the biological self that distinguishes the boundary between "inside" and "outside." Many other creatures have evolved to extend their boundaries in an analogous way. Following Dawkins, Dennett adopts the term "extended phenotype" to describe the extension of boundaries exhibited by, for example, the beaver who builds its dam, the spider who spins its web, the snail that produces its shell, etc. These creatures, obviously, do not understand what they are doing when they extend their phenotypes, nor do they understand what benefits accompany them. Through the process of evolution, many creatures have developed in ways that make "good" border crossings easier and "bad" border crossings more difficult. The human self is no different from these other extended phenotypes, according to Dennett. The human brain extends its phenotype by spinning a self out of words and deeds. And, just like other creatures, the human brain does not have to know what it is doing when it does this. However, a human being's ability to use language marks a sharp contrast between the other creatures of this world and ourselves, not only in how a self is formed, but also in how the self is maintained.

The raw material used to create the extended phenotypes of any non-human creature is found in the external world. For example, the spider spins a web of silk, the snail builds a shell from calcium, the beaver's dam is made from sticks and mud, etc. However, the human environment differs from the "natural" external environment in that it contains "words, words, words," the building blocks of a human self:

> These words are potent elements of our environment that we readily incorporate, ingesting and extruding them, weaving them like spiderwebs into self-protective strings of *narrative*.[74]

[74] Dennett, *Consciousness Explained*, 417.

The self, like the mind, is an artifact. It is created by the brain using meme-vehicles known as words.

To state that the self exists as a narrative structure is to say something quite vague. A narrative is obviously a story, and the narrator the one who tells it. To say that a self is produced through strings of narrative is to say that through the stories we tell others and ourselves, we come to define and represent ourselves as persons, and as persons of a certain kind. As we have seen, the self, like the mind, is an abstraction. The very *source* of self-definition and representation is itself a fiction, one that gets posited as a result of the processes of interpretation, (including self-interpretation). The brain performs a sort of cognitive illusion, wrapping the words of our world around us in such a way that the narratives we incorporate, ingest, and produce, *seem* to point inward towards an author. The self becomes a "center of narrative gravity:"

> These strings or streams of narrative issue forth *as if* from a single source – not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit a *center of narrative gravity*.[75]

Just as the heterophenomenologist attributes beliefs and sensations to a system whose behavior surrounds a certain point between two stereo speakers, so she also attributes a self to any system whose words surround and emanate from a certain gravitational point in the world. Our words "force" others (as a matter of habit, for interpretation can only begin by assuming something is present to be interpreted, i.e. intentionality) to posit a "doer behind the deed," or a "center of narrative gravity." Because the brain is a syntactic engine that can mimic the competence of a semantic engine, here the brain

---

[75] *ibid.*, 418.

mimics the behavior of a Cartesian 'I,' a homunculus, a pontifical neuron, a Central Meaner, or any other concept you prefer to use to describe this phenomenon. This does not entail, for Dennett, that the self can then be identified with the brain, for reasons already outlined in the discussion of the mind. The brain cannot be considered the narrator, since what makes a self what it is are processes of interpretation involving cultural memes (specifically words). Just as a focus on the brain will leave out everything that is important to a mind, so it is also with the self.

According to Dennett, we, as selves, "do not consciously and deliberately figure out what narratives to tell and how to tell them."[76] As human beings, we do not *plan* on doing this, it just happens to us; it is a normal brain function that has developed over thousands of years. Just as enormous termite colonies seem to be the work of a single great artist, but are actually the result of many independently working termites, no individual consciousness spins our "selves;" the words and stories spin us. Therefore, just as there is no such *thing* as *a mind*, there is no such thing as *a self* either; it just seems like there is. As Dennett summarizes:

> A self, according to my theory, is not any old mathematical point, but an abstraction defined by the myriads of attributions and interpretations (including self-attributions and self-interpretations) that have composed the biography of the living body whose Center of Narrative Gravity it is.[77]

The third-person methodology adopted by Dennett throughout his philosophy forces him into this account of the self. Thus, a personal understanding of ourselves *as selves* entails, for Dennett, applying the intentional strategy and the process of heterophenomenology to ourselves for an accurate and reliable description of who we

---

[76] *ibid.*
[77] *ibid.*, 427.

are. On Dennett's view, because it is our brain that is responsible for ingesting and producing strings of narrative, it is also our brain that is responsible for the derivatives of this process, which include the illusions of mind and self. As biological entities we produce and interpret narrative strings as a matter of evolutionary habit. We do not know what we are doing when we do this, nor do we understand the benefits of it; the process *just happens*. The course of evolutionary history has shaped our processes of interpretation and their effects. The fact that we make sense of others and ourselves using the intentional strategy, in which the self and the mind play crucial roles, is a fact of brain evolution on Dennett's theory, and nothing more.

2.3 Re-calibrating Agency

If the self is a fiction, a cognitive illusion, then what becomes of the thought that our selves are self-originating sources of action? That is, if human beings *really* are just physical systems going through the causal motions of absorbing and producing strings of narrative, then must we not conclude that notions such as free will, agency, and responsibility are cognitive illusions? According to Dennett, a self is a "locus of self-control."[78] Yet, if we *really* are physical systems operating in a causal universe – that is, syntactic engines that mimic "mindedness" and selfhood – then we are not *really* in control; it just *seems* like we are. This surely undermines our entire conception of ourselves as autonomous and responsible agents. What has Dennett to say in reply?

One of the true marks of agency is the ability to *make* decisions, to *will*, and not simply be on the receiving end of things happening *to* us. We have the experience of

---

[78] Dennett, Elbow Room: The Varieties of Free Will Worth Wanting (Cambridge: The MIT Press, 1984), 81.

being in control of much of what happens in our lives. The big question for Dennett then, is how do we get a responsible agent capable of making autonomous choices from a causal, non-responsible mechanism? In what way, if any, is an agent the result of acts of self-creation, and not simply the effect of normal brain processes? The short answer for Dennett is that simple decisions and choices are based upon maximizing satisfaction, and our deliberations in such situations eventually lead to an ability to resolve harder choices, including moral decisions. The deliberations we engage with in making these more difficult decisions then create our selves in a special kind of way. Let us examine the details and consequences of this claim.

According to Dennett, much of the "free-will structure" is set before moral considerations even arise. Thus, moral concerns are not *fundamental* to agency. Just as the mind and the self developed gradually through the evolution of the brain, a human agent's ability to engage in the evaluation of her beliefs, desires, intentions, etc. evolves over a lifetime. As we move from childhood to adulthood, we move from "easy meta-level problems of control" to more sophisticated problems that require striking boldly "into the territory of risky, heuristic reasoning."[79] Dennett, again using the analogy of a chess game, argues that children solve their problems with "by the book solutions:" look before you leap, a stitch in time saves nine, don't cry wolf, etc. But once we reach the "middle game" of adulthood, life is much more complicated, requiring agents to move "out of book" in order to make complicated decisions involving our basic projects and goals.[80]

---

[79] *ibid.*, 86.
[80] *ibid.*

According to Dennett, we are always at the mercy of the brain's decision-making processes at a certain level.[81] However, the consideration of deeper problems that often surround one's goals or life projects, as opposed to the resolution of immediately conflicting desires, requires agents to "weigh" various alternatives. In any decision of this kind, Dennett believes that agents must make "relatively blind leaps," but that they can take steps to ensure that they "think the right sorts of things at the right times."[82] Perhaps this means learning from trial and error, learning from our past mistakes, or learning from the mistakes of others, etc. The point is that too much thinking, self-reflection, and self-criticism is overkill; it is clearly irrational "to embark on a limitless round of self-evaluation."[83] One must eventually act, but if we are deciding, for example, between one of two horns in a complicated moral dilemma, the consequences of which may deeply affect our lives, there is "no feasible 'decision procedure'" for determining what the right answer, if there is one, might be. So how do we decide?

How we decide, according to Dennett, involves not only our basic rational decision-making procedures, but also our understanding that who we are as individuals depends upon the decisions we make. Dennett wants to say that by attempting to resolve complex problems, "we define ourselves in the process, by making 'articulate' and more definite that which had been inchoate and ill-formed."[84] Therefore our conceptions of responsibility and morality stem from i) our capacity to make rational decisions, and ii) knowledge that our decisions create who we are.

---

[81] *ibid.*, 87.
[82] *ibid.*
[83] *ibid.*
[84] *ibid.*, 90.

Two objections can be made here. First, it might be complained, as Dennett notes, that moral agency therefore depends a great deal on luck, since one's ability to make rational decisions may be determined by events beyond one's control. The second objection we have seen before. If our brain is responsible for much, if not all, of our rational decision-making structure, then it seems that a causal structure is still in control of our deepest moral considerations, in which case, we cannot *really* be responsible for who we are as agents. This would undermine any sense of autonomy we believe we have in creating ourselves as certain kinds of people. How does Dennett respond?

He argues that the reason we can take credit and responsibility for our rational deliberations is because the ability to deliberate is not a matter of luck at all, but talent. In the course of the evolutionary process, our genes have become "skilled;" they have won contest after contest in the game of genuine prowess throughout the eons. The reason we are here, with an ability to deliberate and control ourselves, is not because our genes "got lucky:"

> ...[S]ince the skills of self-control and deliberation have been put to a fairly severe test over the eons, there is a real basis in fact for our having high expectations about the deliberative skill, and more generally the capacity for self-control, of our fellow human beings. If you weren't very well equipped in that department, you wouldn't have made it to this round of the tournament. Of course some unfortunates, though born of skilled self-controllers, are defective, through no fault of their own. We do not consider them responsible. They are excused. But we do expect a lot from the rest of us, and for good reason. We are not just lucky; we are skilled.[85]

Sure there will be some differences in luck, initial strengths and weaknesses, lucky breaks, etc., but these, Dennett argues, are negligible, once we realize that life is like

---

[85] *ibid.*, 94.

running a marathon and not a hundred yard dash: "luck averages out in the long run."[86] The reason we take responsibility for making the "right decisions" is because we are expected to do so as skilled deliberators, and therefore, we must also take responsibility for when our actions are based upon deliberations that are not-so-skilled.

For Dennett, "there is no incompatibility at all between determinism and deliberation."[87] Expanding upon this claim, he comments on the nature of opportunity, and the idea of what it means to aspire to avoid certain future events. The crux of these considerations depends upon which conception of possibility one is applying in discussions of agency. There are at least three varieties of possibility: i) *logical possibility*, where something is logically possible if it is consistently describable; ii) *physical possibility*, where something is physically possible if it does not violate the laws of physics; and iii) *epistemic possibility*, which concerns possibility as relative to a person's knowledge. Something is epistemically possible for an agent if it is consistent with everything that agent already knows.[88]

To determine which conception of possibility we should apply in our descriptions and explanations of agents, Dennett works backwards and considers whether or not it is possible for evolution, or an artificial intelligence researcher for that matter, to design a deliberator with legitimate "open branches" for the future. He says this is possible. According to Dennett, *any* type of deliberator will have to be equipped with specific ways of partitioning information – specifically a way of recognizing and separating what sets of things can be broken down into an array of *possible* alternatives (raising an

---

[86] *ibid.*, 95.
[87] *ibid.*, 102.
[88] *ibid.*, 147-148.

appendage, turning left or right, etc.), and what sets of things can be broken down into *impossible-to-change* background states (the speed of light, the acceleration force of gravity, etc.). The concept of *possibility* for a system must therefore be an *epistemic* concept, one that regards only "the states of things that are possible-for-all-the-deliberator-knows-or-cares."[89] That is, if x is controllable by the deliberator, then the future is open as far as that system is concerned.

This brings us back to the question of *real* choice versus *apparent* choice. It seems that if we are to be autonomous agents, we need the potential to make *real* choices. Dennett responds to this in two ways. First, according to Dennett, concepts such as "choice," "avoidance," "opportunity," "prevention," "responsibility," etc. are macroscopic and not microscopic concepts:

> All the verbs of "making a difference" involve a tacit comparison between the way the world was *apparently* going to go, and the way it turned out to go. To "change the course of history" is to be the agent whose acts (and before that, whose deliberations) make a *salient* ("pivotal") contribution to the actual trajectory of the world – as judged by contrast with the projected or anticipated (or retrospectively judged) likely trajectory of the world. The use of this concept invokes something like a principle of the inertia of the normal; it relies on the existence of a tacit background of the way things are expected to go "other things being equal," against which an agent's act, being a case of some other thing not being equal, "makes all the difference."[90]

Thus, the human concept of possibility with regard to choice cannot mean *logical* possibility or *physical* possibility, only *epistemic* possibility. On Dennett's view, it does us much philosophical harm to consider these other varieties of possibility in discussions of agency, since a non-epistemic use of possibility requires an objective separation from

---

[89] *ibid.*, 110.
[90] *ibid.*, 125-126.

the everyday background agents find themselves in. Since agents can only make use of the concept of epistemic possibility in their deliberations, this can be the only type of possibility that matters when discussing agency.

One might object that this response favours the first-person perspective of the agent, something Dennett has dismissed from the beginning. By considering only epistemic possibility in discussions of agency, do we not legitimate how the world seems to the deliberator? Recall however, that Dennett's promotion of the third-person perspective is not intended to tell us "what it is like" to be a human being. It is intended to be a neutral method for collecting data. With science we have found an alternative way to describe our place in the universe, but we cannot use it to describe ourselves as *agents*, since agency and action are macroscopic terms. According to Dennett:

> The manifest image of human beings is the everyday world of colored things (not swarms of molecules); our eyes are macroscopes, not microscopes. We also perceive only "middle-sized" rates of change; things that happen faster or slower are imperceptible to our naked eyes and ears.[91]

We lose the macroscopic concepts associated with agency if we move to an alternative perspective (either the design or physical stance). The macroscopic concept of action is what forces Dennett to articulate the intentional strategy in the first place, and therefore, the adoption of epistemic possibility to account for agency does not conflict with the third-person perspective. Just as we cannot help but recognize intentional patterns, we cannot help but act *as if* we have an open future. Heterophenomenology therefore dictates that we use the concept of epistemic possibility in our construction of notional worlds, since this is the phenomenal space the subject obviously "believes" in. It is

---

[91] *ibid.*, 114.

helpful to think of the intentional strategy and heterophenomenology as an explanation of action from the third-person perspective using macroscopic concepts as tools.

Does this mean that in order to think of ourselves as agents, we must keep ourselves under the *illusion* of agency? Dennett asks us to imagine a robot that is designed for a specific mission. It is designed to make the most of its opportunities, which simply means that it can be described as having *interests*, and therefore *beliefs* and *desires*, in a way that is required by the intentional stance. A robot has a *robot-opportunity* "whenever it is caused by such special-interest events to "consider," and if it so decides, to plan and execute, a timely project designed by it to further its interests in the light of those events."[92] This simply derives from how we formulate our general theories of interpretation according to the intentional strategy. Here, we decide to treat the robot whose behaviour is to be predicted as a rational agent; then we figure out what beliefs the robot ought to have, given its place in the world and its purpose. Then we figure out what desires it ought to have, on the same considerations, and finally we predict that this rational robot will act to further its interests in the light of its beliefs.

In order for the robot to perform certain tasks, it must be programmed to follow specific rules. Thus, the robot is *determined* – the program is the cause of all of its "choices" and "actions." The crucial objection turns on just this point. Surely, it will be argued that while the robot can be *described* as having opportunities, choices, and the ability to perform actions, only human beings can be considered to have the *true* marks of agency – *real* opportunities and *real* choices, not just the *appearance* of them. This objection recalls one we have heard before. It was argued that while other systems can be

---

[92] *ibid.*, 116.

described as having beliefs, only human beings can be considered to be the *true believers*. Dennett's response here is exactly the same as it was previously. Just as folk psychology explains *action* in terms of the intentional states of the agent, the everyday explanation of *agency* must rest on a similar conception of "folk-autonomy," where the explanation of opportunity and choice is made in terms of epistemic possibility, i.e. choice "as it seems to the agent." If opportunity and choice are to be anything, on Dennett's view, they must have this deflated meaning. Dennett wants to think of human choices and opportunities as being exactly like that of our designed robot.

Thus, on Dennett's view, responsibility is not exactly what we thought it was or would expect it to be: "If we are to be found responsible at all, it will have to be a modest, naturalized, slightly diminished responsibility."[93] We cannot escape the fact that we live in a deterministic universe, and therefore our sense of responsibility must accommodate this fact. Consider:

> Knowing that I will always be somewhat at the mercy of the considerations that merely happen to occur to me as time rushes on, knowing that I cannot entirely control this process of deliberation, I may take steps to bias the likelihood of certain sorts of considerations routinely "coming to mind" in certain critical situations. For instance, I might try to cultivate the habit of counting to ten in my mind before saying anything at all about Ronald Reagan, having learned that the deliberation time thus gained pays off handsomely in cutting down regrettable outbreaks of intemperate commentary. ... The main thing is to see to it that I will jolly well do otherwise in similar situations in the future.[94]

On Dennett's view, being responsible means taking it upon yourself to avoid certain future outcomes.

---

[93] *ibid.*, 158.
[94] *ibid.*, 143.

But what is our *motivation* for being responsible, good, moral, etc.? For Dennett, it is simple: "*[I]t is rational* for us to esteem free will and covet responsibility."[95] Thus, not only is *being* responsible acting out of "rational well-designedness," but our motivation for *taking* responsibility is based upon rational considerations as well:

> Why then do we want to hold people – ourselves included – responsible? "By holding someone responsible and acting accordingly, we may cause him to shed an undesirable trait, and this is useful regardless of whether that trait is of his making." (Gomberg 1978, p. 208) Once again, the utility of a certain measure of arbitrariness is made visible. ... we are rewarded for adopting this strategy by the higher proportion of "responsible" behavior we thereby inculcate.
> ... By somewhat arbitrarily holding people responsible for their actions, and making sure they realize that they will be held responsible, we constrain the risk-taking in the design (and redesign) of their characters within tolerable bounds.[96]

It is in our best interest to be responsible. If we act responsibly, then not only do we avoid making embarrassing mistakes, we also avoid being punished by our community for the irresponsible actions that frequently lead to the harm of others.

Here we can tie together Dennett's positive account of self and agency and its consequences for his view for self-understanding. Given the third-person approach towards the explanation of action and agency, we saw Dennett proceed with an evolutionary account of how the brain produces the cognitive illusion of mind and the fiction of self-hood. This forced Dennett to re-evaluate the notions of agency – autonomy, responsibility, and morality. However, because Dennett focuses almost exclusively on the evolution of the brain, which is a physical system causally determined in a physical universe, the third-person perspective that attempts to make new sense of the macroscopic world of human beings must still cling to the deterministic physical

---

[95] *ibid.*, 155.
[96] *ibid.*, 163-165.

world. By keeping determinism in play, Dennett must deflate our moral concepts to fit this scheme. Some problems with this method for understanding human agency will now be addressed, and in chapters 3-4 we will see why an alternative approach is superior.

2.4 Problems with Dennett's Account

We can make two strong objections against Dennett, with the underlying premise behind each being a criticism of his methodology as an adequate approach to the subject of action and agency. I want to argue that the account of self and agency that derives from Dennett's theory of action interpretation undermines his entire position since i) we are left with a view of the self in which we cannot recognize ourselves, and ii) the ethical dimension associated with Dennett's account of agency is deflationary to a fault. While these two objections are related, they warrant separate discussion. I shall examine each in turn before considering a more convincing account of action and agency.

The first objection concerns the plausibility of seeing ourselves as "centers of narrative gravity." By defining the self as a center of narrative gravity, Dennett seems to have oversimplified the relationship between agents and interpreters. Recall that as centers of narrative gravity, our selves are abstractions *defined* by the myriads of attributions and interpretations (including self-attributions and self-interpretations) that compose the *biography* of the living body whose center of narrative gravity it is. However, this is not *all* a self is defined by, and for this reason, the concept of a self as a center of narrative gravity is one-dimensional.

The content of our biographies must be distinguished from the narrator of that content. The subject matter that composes a biography of a self does not reveal what is most important to us as selves; the most important distinctions between selves can be

found only by looking at *the telling* of the biography. Deborah Knight, taking issue with this feature of Dennett's thought, writes:

> [T]he self can be produced in the telling because it is the telling of the tale that matters, regardless of the ostensible content of the tale as told. In this way, the tales I tell about others are also important in constituting myself, even though I may not figure at all as an agent within the tale, and may not refer to myself in the first-person at all in the recounting. It is not the content of the stories we tell, but the telling, that constitutes one self as opposed to another. A's gossip about B may be far more revealing about, for example, A's envy, than it is about B.[97]

Dennett seems to miss an important role for any interpreter: to recognize the speaker as the teller of the tale. This is not an aspect that is simply overlooked by Dennett. The omission follows necessarily from his third-person methodology. As we have seen, Dennett's project consists, in part, in debunking any theory of mind or action that makes reference to the mind or the self as entities or things. These are philosophical errors, according to Dennett, rooted in the first-person perspective. Thus, the third-person perspective relegates the mind and self to the level of abstraction. However, if there is no such *thing* as a mind or self, then there can be no such *thing* as a *teller of the tale* either. But clearly, as Knight argues in the quote above, who we are as persons depends heavily upon our interpreters for self-definition, not solely for reasons of content, but also for reasons of recognition. An interpreter's recognition of herself *as a teller*, regardless of whether "teller" refers to an independently existing entity or not, can and does define us in a way that is essential to understanding ourselves as persons.

This leads to a stronger point. Since we need to understand ourselves as occupying the position of "teller," we cannot see ourselves as wholly constituted by stories told about us. Recall that for Dennett no individual consciousness spins our "selves;" rather,

---

[97] Deborah Knight, "Selves, Interpreters, Narrators," *Philosophy and Literature*, 1994, 18: 281.

the words and stories *spin us*. If, as Knight argues, we need to understand ourselves as occupying the position of "teller," then we are recognizing something prior to and not wholly determined by the stories that are told. Dennett cannot account for this because, on his view, the self is a fiction, a cognitive illusion – there is no*thing* prior to the stories told. But our everyday practices of interpretation contradict this claim. Dennett's methodology is unable to account for the entire process of self-definition, as well as the facets of agency that surround our conception of ourselves as *tellers*. As Knight concludes:

> Dennett offers no insight into the poetical activity which is central to the narrativization of an agent's life, nor any sense of how the basic cognitive abilities that characterize intentional agents support the interpretive activities directed toward trying to make sense of such agents.[98]

Knight argues that if anyone's biography is told from an omniscient third-person perspective – which seems to be what Dennett recommends – then the content of that biography must include *everything*. The purpose of such a third-person perspective is to collect data in a neutral manner, and therefore, nothing can be omitted since to leave some things out and others in would require a reason for doing so, which would be biased on any consideration. But Knight makes the point that who we are as persons is not, and cannot be, the *whole* story:

> [W]hat we long for is not the Whole Unabridged Story; it is the Whole Unabridged Story of *Me*. *That* story, of whatever interest to the subject in question, would only be an insignificant bit of background detail in the Whole Unabridged Story. The Whole Unabridged Story is not what we were looking for; we wanted our own volume devoted to ourselves.[99]

Who we are is what *matters* to *us*. But while this is only a negligible part of the Whole

---

[98] *ibid.*, 284.
[99] *ibid.*, 279.

Unabridged Story, it is the most important part in coming to understand who we are as persons. We cannot incorporate this fact into Dennett's account since there is no room in an evolutionary story for *mattering*. Thus, attempting to understand the self by focussing on narrative does not necessarily render a complete understanding of the self, especially when the focus of the narrative concerns only biographical content. What matters in a narrative of the self, as we have seen, is not merely the other characters involved in the telling, but the teller herself and what matters to her. It is authorship that we are most concerned with in trying to understand ourselves. We want our own story because this is how our life seems for us, and this is what an understanding of the self should entail.

If what is important to being a self is the telling of our own story, then this gives us the elbow room required for the kind of self-creation and autonomy we are all familiar with, yet it is the same kind that cannot find its way into Dennett's account. According to Dennett, there is no "I" to weave a self, and therefore by definition there cannot be an "I" to spin our own narrative; as we observed, the words and stories of the world spin us. For Dennett, this gives us the *appearance* of a self, but not an independently existing entity. Although it might *seem* to us that we are the self-originating sources of our actions, lying outside the causal stream with our special kind of autonomy, we cannot escape the fact that we are always at the mercy of our brain's decision-making structure. Thus, as we have seen, autonomy and responsibility are present in Dennett's account, just not in the form we are most familiar with.

But wait a minute. From the beginning, Dennett's approach to action and agency has been anti-reductionist. The point of the intentional stance was to be able to speak about the mind without having to speak about the brain. This is because "mind" is

associated with the category of terms we use to make sense of ourselves, in particular, the macroscopic concepts of "action" and "agency." However, if a theory of agency and action is being based upon how we actually go about making sense of ourselves (using macroscopic concepts), then should we not have an account that makes sense to us, as the agents the theory is supposed to be about?

For example, recall Dennett's account of responsibility. We saw that for Dennett, being responsible means taking it upon oneself to avoid certain future outcomes. We count to ten before saying anything about Ronald Reagan, because a forced deliberation like this prevents regrettable outbursts. But why is the outbreak regrettable? Perhaps I may offend a strong Reagan supporter, or perhaps a loud outburst will convince others that I am arrogant and pompous. It would be a good thing if these were avoided in the future, and so I count to ten.

But why should I care about the future? Isn't it because it is *me*, and therefore, *my* future? Isn't it because *I* will be the one experiencing embarrassment, humiliation, shame, guilt, etc. upon making an ill-timed comment about Reagan? Not only do I not want to feel these negative emotions, but I also do not want others thinking of me as conceited, immodest, and smug. Dennett cannot address this aspect of self-understanding. With the self as an abstraction, the focus must inevitably turn towards action, and not what matters to the agents themselves. By omitting our experience as a self, as agents with concerns, we lose what is central to our experience as human beings: *That life is happening to me*! There is no room in Dennett's theory for me as the teller of my tale, for my *me-ness*. We are therefore left with a theory that does not make sense of *us* at all. By relying on an evolutionary theory of the brain in order to explain human

action and agency, Dennett – who wants to stay in the world of the macroscopic – leaves out the most important macroscopic concept: *Our experience of ourselves as minded beings*. It is this *experience* that we want to explain. Not how it came to be this way (which is a scientific/evolutionary explanation), but how we should explain ourselves to each other in order to understand ourselves *as the beings we seem to be*. If we cannot see *ourselves* in Dennett's account, surely this must undermine any complete understanding or explanation of *agency* that he envisioned.

The second main objection against Dennett stems from the ethical dimension of his account of agency. Recall that for Dennett our ethical deliberations follow from a pattern of rational deliberations, where "rationality" is defined as "getting ahead cognitively in the world." But let us examine how this conception fares in some simple examples, such as weakness of the will. Consider acts of infidelity, betrayal of a loved one, stealing, or something similar. On Dennett's account, since ethical deliberations follow from rational deliberations, how can he explain cases where an agent acts immorally and irrationally (e.g. they commit adultery knowing that they have more reason to be faithful)? I will argue that Dennett cannot properly account for these cases. This makes his conception of morality deflationary to a fault.

First, recall that being responsible is identified with being a skilled deliberator, for Dennett. The reason we take responsibility for making the "right" decisions is because we are expected to, as skilled deliberators. Consider Dennett's comparison between himself and some "tough young hoodlum:"

> The day someone happens to make the mistake of insulting [the young tough hoodlum] is the day his life of violent crime begins. If the sort of temptation that would turn him into a murderer were to flash before my eyes, however, I would almost certainly resist it, I would have no one but

> myself to blame. I am supposed to be good at resisting such temptations – and in fact most of them are child's play. I'm so good at them, *I don't even notice them as opportunities*. Consider fine chess players, who never even notice the stupid move opportunities. It is not just luck that keeps them from making the "patzer" plays.[100]

As a skilled deliberator, Dennett is quite good at staying out of trouble. In fact, he is so good that he does not even recognize certain temptations as opportunities. Does the skilled deliberator always avoid weakness of will?

An agent has a weak will if he acts intentionally against his own best judgement, that is, against what he knows or believes to be a better course of action. Consider one of Donald Davidson's descriptions of the experience of incontinence:

> The image we get of incontinence from Aristotle, Aquinas, and Hare is a battle or a struggle between two contestants. Each contestant is armed with his argument or principle. One side may be labelled 'passion' and the other 'reason'; they fight; one side wins, the wrong side, the side called 'passion (or 'lust' or 'pleasure'). ... Here there are three actors on the stage: reason, desire, and the one who lets desire get the upper hand. The third actor is perhaps named 'The Will' (or 'Conscience'). It is up to The Will to decide who wins the battle. If The Will is strong, he gives the palm to reason; if he is weak, he may allow pleasure and passion the upper hand.[101]

As skilled deliberators, we are supposed to choose reason, what we *ought* to do, over impulse or pleasure. This is what keeps us out of trouble, and prevents us from acting wrongly or immorally. But how does Dennett explain when our deliberations fail, that is, when we succumb to desire against our own better judgement?

Recall that the intentional stance will have difficulty explaining and describing errors. We saw that the attempt to describe and explain the mistake of the lemonade seller using folk psychological terms fails, because it is an attempt to deliver a

---

[100] *ibid.*, 99.
[101] Donald Davidson, "How is Weakness of the Will Possible?" *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 35.

rationalizing explanation of an error, which, by definition, precludes any rational understanding. Thus, it was concluded that we must descend from the level of beliefs and desires to some other level of theory to explain mistakes. It seems that this is therefore a possible explanation for cases of weakness of the will. Since incontinence by definition means going against reason – one's better judgement – a weak will can be explained by descending beneath an explanatory scheme of rationality. Therefore, as it is with mistakes, explaining incontinence requires switching to either the physical or design stance. Thus, for Dennett, any immoral behaviour resulting from weakness of the will has nothing to do with our state of *mind*, but rather our state of *brain*, since this is where the physical or design stance takes us. On this view, being incontinent, like making a mistake, must be the result of a "system failure."

However, instances of incontinence must be distinguished from the occurrence of pure errors. Mistakes, like giving incorrect change, are errors of reasoning that occur beneath our level of awareness, but acts of incontinence are performed *intentionally*. We know what we are doing when we reach for that second doughnut while on a diet, and the man who has intercourse with his best friend's wife knows it is wrong to do so, but does it anyway. Explaining these actions as "system failures" does not work for weakness of the will, since we choose, intentionally, to do something that we know is wrong or irrational. Thus, it seems we have an example of an intentional action that does not fit Dennett's explanatory scheme, since we have an example of an intentional, irrational action.

Dennett disagrees. According to Dennett, incontinence is a "universal pathology," an inescapable case of "local fatalism." In such cases, deliberation is futile, but our

course of action is not necessarily determined, it is just "informationally insensitive."[102] Since deliberation is *generally* effective, Dennett believes we can ignore cases of weakness of the will, which are just instances where deliberating does not make a difference. We are simply lucky that such occurrences are "abnormal to our world."[103]

Basically, Dennett is denying the possibility of an intentional, irrational action. If what we observe looks to be like such a case, we are mistaken. For an action to be considered intentional, that is, capable of receiving a description using "belief" and "desire," it must at least be possible to act on those beliefs and desires; it must be possible to describe an agent as having beliefs and desires that *caused* him to act. But with incontinence, we do not have this situation. To use Dennett's example, a man who is "madly and helplessly in love" cannot use reason to convince himself otherwise "since there is a compensating mechanism at work that will convert any supposed blemish into a virtue: she's not stupid, she's down-to-earth; she's not selfish, she's spirited."[104] Somehow, the deliberation process is being removed from the causal order of things, and therefore, weakness of the will cannot elicit an intentional description.

It is Dennett's contention then, that an act of incontinence cannot be described using the intentional stance because the action cannot be described as being rationalized by a certain set of beliefs and desires. But then, how is that action to be described? According to Dennett's theory, we must descend from the level of beliefs and desires to another level of theory (the physical or design stance) in order to obtain the proper description. However, the physical and design stances operate from a causal perspective,

---

[102] Dennett, *Elbow Room*, 105.
[103] *ibid.*, 106.
[104] *ibid.*, 105.

not an intentional one. If true, then it does not make any sense to call acts of incontinence "intentional," since they cannot be described from the intentional stance. Incontinent acts must therefore be caused, but does this not bring us back to describing these acts as causally determined, and therefore as a system failure? It seems that Dennett must concede that acts of incontinence are caused at the level of the brain, since there is no intentional description that makes any sense. Perhaps, since our rational decision making procedures seem to be disconnected from the part of the brain that moves us to action, acts of incontinence are the result of a "system failure," although of a slightly different sort than pure errors.

Recall Dennett's own statement that if the sort of temptation that would turn a young tough hoodlum into a murderer were to flash before his eyes, he, Dennett, would almost certainly resist it. Therefore, being moral in this instance is the result of being a good deliberator. But if not resisting such a temptation is considered weakness of the will, then Dennett's claim is not that a *poor* deliberation occurred, but that *no* deliberation occurred. Dennett cannot claim that a poor deliberation occurred because cases of incontinence seem to involve intentional actions. But on Dennett's view, intentional actions are assumed rational as a function of the intentional strategy. Therefore, the idea that an agent would *choose* to do something against his better judgement – against what he knows to be a better decision – is not a poor decision, but a decision that makes no sense because it is by definition irrational, and cannot be given a rationalizing explanation (with beliefs and desires). Therefore, not resisting temptation becomes the result of a causal *faux pas* within the brain.

However, if not resisting the temptation is a part of the deterministic universe, then why not also our resisting temptation? It seems that if we resist temptation on Dennett's view, then we were deliberating, but if we do not, then we were not deliberating. This is not absurd on Dennett's view, since what is required for autonomy is only an *epistemic possibility* between choices. If the alternative in a case of weakness of the will is between resisting and not resisting temptation, then even though there is only one future branch that is *really* open, this counts as a legitimate choice for Dennett, because we are only concerned with possibility for-all-the-agent-knows-or-cares.

This cannot be right however. If an agent resists temptation, the action gets described as making a choice, and the agent receives accolades for being a "strong" and "morally responsible" person. If the agent does not resist temptation however, the action is described as a system failure, and the agent does not get described as "weak," but as a victim of causal processes that were beyond his control. But it makes no sense to claim that the choice of an agent is based upon *epistemic* possibility, and then in circumstances of weakness of the will, describe that choice on a level of theory focussing on *physical* possibility. What is the point of trying to approach a theory of the self using macroscopic concepts when you must continually purge these concepts (first for errors, and now, incontinence) for a complete understanding of agency? It seems a consistent theory of agency requires that we explain our moral behaviour based upon a sense of self that is necessarily irrational at times, which is impossible on Dennett's account.

In chapters 3-4 we will examine the work of Charles Taylor. Taylor gives an explanation of agency and action that accounts for our experience as minded human beings in a way that is able to respond to the objections raised here.

Chapter 3

Taylor's Inescapable Horizons: Moral Agency and Self-Interpretation

Charles Taylor's account of the self stems from his interest in self-interpretation. Taylor maintains that who we are as selves – and this includes our abilities to think and act – is socially defined: our personal identities are the result of processes of interpretation. Following Herder, Nietzsche, and Heidegger, Taylor believes that we are required to treat the self as situated in the world. Not only must an understanding of the world begin with the recognition that knowledge is pursued in light of human goals and needs, but an understanding of the self requires that we see ourselves as inseparable from the world.

This chapter outlines Taylor's hermeneutic approach to the self, and explicates his view of the connection between self-interpretation, ethical reflection, self-identity, and responsibility. If Taylor's analysis is correct, this has important consequences for Dennett's methodology. In Chapter Four, I hope to show in more detail why Dennett's understanding of the self is flawed, specifically why the methodology Dennett employs inevitably undermines his entire account. I will argue that Taylor's approach encompasses precisely what is required in any understanding of human agency. His position offers us a more plausible understanding of the self, which accounts for human experience in a way that cannot be captured by Dennett's third-personal approach.

3.1 Important Distinctions

We can trace the development of Taylor's account of the self back to an important distinction he deploys between first-order and second-order desires. Having a first-order desire means simply desiring some or other object or outcome (I want to eat an ice cream), whereas a second-order desire has as its object a specific first-order desire (I want to stop desiring to eat ice cream). Following Harry Frankfurt, who originally marked this distinction, Taylor believes that the formation of second-order desires is unique to human beings, since it involves an *evaluation* of first-order desires, something which requires self-reflection.

However, to understand what is unique to *agency*, Taylor argues, requires that we make a further distinction between forms of evaluation. For Taylor, reflective self-evaluations can be further distinguished as either weak or strong.[105] Weak evaluation involves a person weighing two desires in light of preferences or interests that are basically self-interested, pragmatic and/or utilitarian in nature, such as convenience, pleasure, etc. For example, I might decide between chocolate and strawberry ice cream by considering what I am "in the mood" for. Or I may decide to take the boring but efficient four-lane highway to the cottage, even though an alternative route passes through beautiful Algonquin Park, since saving time is more important to me than aesthetic appreciation.

In contrast, if a person is engaged in a strong evaluation, then the motivation for

---

[105] Charles Taylor, "What Is Human Agency?" *Human Agency and Language* (Cambridge: Cambridge University Press, 1985), 16.

choice contains a consideration of quality. Our actions and desires are evaluated as "high or low," "noble or base," "good or bad," etc. To say that the motivation for a first-order desire by a strong evaluation is qualitative is to say that it is not simply motivated by the *satisfaction* of that first-order desire. For example, my desire for ice cream may be motivated simply by my desire for something sweet, and therefore, my motivation for eating the ice cream is the satisfaction of my craving. However, if I try to refrain from acting on my jealousy because I consider such an act to be "ignoble" or "unworthy," then I have been motivated by a consideration of quality, and not simply a craving. I refrain from acting out of jealousy because I do not want to be the kind of person who does that sort of thing.

Of course, in the above example, we can describe the choice of route to the cottage as "good," but here our evaluation is simply linked to the satisfaction of a specific outcome, and does not involve what Taylor calls "qualitative distinctions." The motivation behind the choice of route is concerned only with the idea of saving time. We want to get to the cottage quickly, because we want to have more time to relax, swim, golf, etc. This reason, however, is simply pragmatic. But if we are motivated to take the efficient route to the cottage because we abhor anything associated with inefficiency and sloth, then our choice of route can be described as a "good" one in a strong evaluative sense, because it is being contrasted with a "poorer" form of life, and here there is a consideration of quality.

Decisions based upon strong evaluations are described as "noble" or "base," "good" or "bad," "worthy" or "unworthy," etc., because the motivation behind the desire is *itself* considered valuable or value-less. According to Taylor, the desire for each choice in a

strong evaluation contains an intrinsic conception of *worth*.[106] So while both weak and strong evaluations involve deliberating between desires, in strong evaluations we are considering the *worth* of a choice; our primary concern is not whether it is efficient, pleasurable, etc.

Taylor observes that strong evaluations have two notable properties. The first is that in order for the choice to be described using evaluative terms of worth, such as "virtuous" or "noble," that choice must be *more* than simply *desired* by an agent. Note the difference in being motivated to choose something because it tastes good, or saves time, for instance, and being moved towards something because it is "noble," "virtuous," "higher," etc. There is no intrinsic worth found within the choice of chocolate ice cream itself, because the motivation behind the choice is the satisfaction of a personal pleasure. But performing a courageous act, for example, is not motivated in the same way, even if one finds pleasure in acting courageously. There is already a conception of worth attached to the terms that are used to describe acts of courage, because the motivation behind acting courageously is seen as worthy, whether one's personal tastes coincide or not. This is why it is possible to judge some desires, like the desire to overeat, as "wrong," "unworthy," or "superficial," etc., even though someone may find it enjoyable.[107] If *worthiness* were a function of desire *only*, this condemnation would not make any sense.

Taylor notes a second difference between strong and weak evaluations. This concerns how we are to understand the motivations lying behind each type of evaluation. According to Taylor, since there is a conception of worth motivating desires in strong

---

[106] Taylor, "What Is Human Agency?", 16.
[107] *ibid.*, 18.

evaluations, the alternative set aside in any strong evaluation cannot therefore be based simply upon a contingent incompatibility with another alternative. For example, in a weak evaluation, my choice between chocolate and strawberry ice cream could as easily have been a choice between chocolate and vanilla ice cream, between having chocolate ice cream or going to the movies, or between eating chocolate ice cream now as opposed to later. My choice between any of these two options is going to be motivated by how I *feel* at a given time, and therefore, there is no *necessary* contrast between the two options. Considering how it may be possible to do *both* on a separate occasion shows that a contrast of desirables in a weak evaluation is contingent.

However, in a strong evaluation there is a necessary contrast of desirables. If I desire to act courageously, then I necessarily desire not to act in a cowardly way. If I describe a courageous act as "noble," for example, then acting out of cowardice in the same situation must warrant the description of "ignoble." However, my desire to eat chocolate ice cream does not entail that I do not also desire to eat strawberry ice cream, or do not desire to eat vanilla ice cream later, etc. Likewise, if my choice is between eating chocolate ice cream or going to the movies, this does not prevent me at some other time from *both* eating chocolate ice cream *and* going to the movies. One does not *have* to contrast chocolate ice cream with strawberry ice cream, or eating ice cream now as opposed to later, although one *can* contrast them. In strong evaluation, conversely, there is a necessary tension between desirables.

The two properties identified with strong evaluations mentioned above – the motivation for choice is not a simple function of desire, and a necessary contrast exists between desirables – do not apply to weak evaluations because of the conception of

worth associated with the former. Due to this difference of worth, Taylor wants to say that the aspiration to eat chocolate ice cream, or the desire to take a shorter route to the cottage, are different in an important way from the aspirations to be courageous, noble, good, etc. The distinction does not rest so much upon quantitative or qualitative differences between the *alternatives* in each type of evaluation. Rather, the more important distinction for Taylor is between two different kinds of *evaluating*, where the *motivation* for one set of desires is based upon considerations of worth, while the other is not.[108] With strong evaluations, Taylor wants to say that there is a *deep* conflict of desirables, and this has important consequences for any agent who engages with this type of struggle.

There is a difference, for Taylor, between the type of self that engages in strong evaluations (which he calls a "strong weigher") and the kind of self that is involved in weak evaluations ("weak" or "simple weigher").[109] According to Taylor, simple weighers possess what is necessary to be a person: reflection, evaluation, and a will; but what they lack is a certain *depth* that is connected to strong evaluators. A strong weigher has a rich language for self-description, a vocabulary of worth[110] (which includes terms such as "noble," "base," "high," "low," etc.). If we consider Taylor's example of the compulsive overeater, we can see the importance of this. Consider a scenario where a man is addicted to over-eating, such that he cannot resist rich desserts. Now by employing Taylor's distinction, the man's evaluation of the situation can take one of two forms: that of a strong weigher, or that of a weak weigher. If a strong weigher reflects

---

[108] *ibid.*, 19.
[109] *ibid.*, 23.
[110] *ibid.*, 24.

upon the situation, he may come to realize that moderation in eating is the key, since "someone who has so little control over his appetites that he would let his health go to pot over cream-cake is not an admirable person."[111] Thus, he may aspire to be someone who wants their bodily appetites to serve higher aspirations, and not simply be the cause of his ever-increasing degradation. Thus, we have a strong weigher's *motivation* for not eating the rich desserts.

However, a weak weigher will consider this situation in an entirely different manner. He may, for instance, simply consider the facts: "By eating too many rich desserts, I increase my cholesterol intake, which leads to an increased chance of heart disease or kidney failure." Or perhaps he will consider some other benefits of restraint: "If I continually break down and eat these rich desserts, I will eventually get fat, and this will have a detrimental effect on my golf swing, as well as my ability to attract members of the opposite sex. Therefore, I should pass over the dessert tray if I want to continuously break ninety on the golf course, and get a date." Thus we have a weak weigher's *motivation* for not eating the rich desserts.

So what we have with this example of the overeater is a contrast of self-interpretations. The obvious question arises as to which is more valid – that is, which is more "faithful to reality."[112] After all, either interpretation of the situation can be responsible for the man refraining from indulging in the rich desserts. The only difference seems to be the language he uses in describing the different motivations behind his restraint, but is this really a significant difference? Perhaps we can do without one of these self-interpretations, or reduce one interpretation to the other.

---

[111] *ibid.*, 21.
[112] *ibid.*, 22.

We need to consider the importance of the strong weigher's access to a rich language of self-description. As we saw, Taylor believes that weak weighers lack a certain *depth*, but what exactly is the significance of this? According to Taylor, the difference a vocabulary of worth makes is seen as an ability to answer certain types of questions. The strong weigher, because he is using a vocabulary of worth in which the terms are contrastively described, is able to articulate actually *why* he chooses one desire over another, that is, *why* the choice of restraint is superior to the choice of indulging in rich desserts.

According to Taylor, "the reflection of the simple weigher terminates in the inarticulate experience that A is more attractive than B."[113] Taylor argues that the evaluation is inarticulate because the motivation for choosing A over B, for example, is directed by weighing how each choice *feels*. Since evaluating A and B is a weak evaluation, there is no necessary contrast between the two options, and therefore, all one can say is that "I prefer, at this time, A over B because I like it." However, the strong weigher can say *more* with a language that is contrastively described:

> I cannot tell you perhaps very volubly why Bach is greater than Liszt, say, but I am not totally inarticulate: I can speak of the 'depth' of Bach, for instance, a word one only understands against a corresponding use of 'shallow', which, unfortunately, applies to Liszt. In this regard I am way ahead of where I am in articulating why I now prefer that éclair to the mille feuilles; about this I can say nothing (not even that it tastes better, which I could say, for instance, in explaining my preference for éclairs over brussels-sprouts; but even this is on the verge of inarticulacy...[114]

The choice of A over B for a weak weigher is motivated only by how A *feels* compared with B. But this is also the motivation the weak weigher has towards his choice of A over C, A over D, B over D, etc. Thus, a motivation based upon how one

---

[113] *ibid.*, 24.

feels delivers no significant information about why one particular choice is superior to another, and therefore, the choice seems to be arbitrary. But because strong evaluations involve terms that are contrastively described, understanding one term entails an understanding of its contrast. If someone is described to me as noble, I have access to a certain amount of information because "noble" is a part of a rich vocabulary of evaluative terms. Using such a "thick" term entails that a judgement of worth has been made based upon some external criteria, and therefore, the evaluation is not only articulable, but also intelligible to others. So a vocabulary of worth not only allows someone to be more articulate (because *why* A is preferred over B is contained within the evaluative term, since we know the contrary), it also allows for *deeper* insights (since the motivation for the choice of A over B is a consideration of *worth*, and not the mere satisfaction of a desire).

Thus, the first important difference between a strong weigher and a weak weigher is that the strong weigher uses a vocabulary of worth, and thereby has access to articulable considerations at a *deep* level. This leads to a second difference between strong and weak weighers: the quality of life being considered in the evaluation. When someone eschews a cowardly act, for example, they do so because they desire to be a certain kind of human being (i.e. a noble one). However, the weak weigher cannot attain the kind of depth attached to this mode of existence. Choosing an éclair over a mille feuilles, for example, does not impact upon a human life in any deep way, because the motivation behind it is one that involves only short-term gains, not forms of life:

> ...a reflection about what we feel like more, which is all a simple weigher can do in assessing motivations, keeps us as it were at the periphery; a reflection on the kind of being we are takes us to the centre of our

---

[114] *ibid.*, 25.

> existence as agents. Strong evaluation is not just a condition of articulacy about preferences, but also about the quality of life, the kind of beings we are or want to be.[115]

The reason Taylor believes strong evaluations take us to the centre of who we are as agents is related to the idea that second-order evaluations of first-order desires are thought to be what is unique to human beings. Therefore, if strong evaluations involve a certain depth, then by engaging with them, we go deeper into what is unique to us as human beings. It seems that reflection upon strong evaluations allow us to explore our humanity in a way weak evaluations cannot. Therefore, what concerns Taylor is not which self-interpretations are more "faithful to reality" per se, but rather, which self-interpretations involve a certain quality or depth.

3.2 Moral Frameworks

In the previous section it was noted that in order to apply qualitative judgements of worth to a choice (such as "noble," "higher," etc.), that choice must be more than simply desired by the agent. As Taylor notes, "I couldn't just *decide* that the most significant action is wiggling my toes in warm mud. Without a special explanation, this is not an intelligible claim."[116] Therefore, something *else* is required to confer worth on a choice. What is this *something else*? Without it we cannot completely understand strong evaluations and the type of agent who engages with them.

Something else is required for the use of strong evaluative terms because feeling a certain way about something can in no way confer significance or importance onto the "object" that feeling is directed towards. Our value terms would lose all meaning if this

---

[115] *ibid.*, 26.
[116] Taylor, *The Malaise of Modernity*, 36.

were possible, since *everything* could then be valuable, and therefore, *nothing* would be.[117] Having the freedom to choose is obviously important, but unless some choices are more significant then others, choice becomes arbitrary. To decide that what *you* feel is most important is to collapse what Taylor calls "horizons of significance."[118] Horizons of significance are backgrounds of intelligibility in light of which things take on importance. More fundamentally though, horizons of significance are *given*; they are grounded in our histories and past connections, things we have little or no control over. Therefore, Taylor's claim is that there can be no consideration of worth without an orientation towards or within, a certain framework of reference.

These horizons of significance – or "moral frameworks" as Taylor sometimes refers to them – confer worth on the choices we make in strong evaluations. The framework endows qualities to the options in a strong evaluation, and because of these qualities we are motivated to choose them. To clarify what having a framework means, let us consider some examples. Past frameworks include the "warrior honour ethic," where it was fame and glory that marked the "high life." Another is Plato's virtue ethic, where the "good life" was given by a method of reason defined by a vision of order in the cosmos.[119] Taylor believes that much of our modern framework consists in the "affirmation of the ordinary life."[120] In this framework, the moral or "good life" is marked by denouncing previous forms of life and the ethical views which accompanied them for their implied elitism. Thus, value and meaning are found, not "outside" the ordinary life of production and reproduction (i.e. in God or the Cosmos), but in different

---

[117] *ibid.*, 39
[118] *ibid.*
[119] Taylor, *Sources of the Self: The Making of the Modern Identity* (Cambridge: Harvard University Press, 1989), 20.

manifestations of the ordinary life. Current versions of this include Utilitarianism, which finds value in pleasure and instrumental rationality, and Marxism, which locates value in free self-creation and enlightened rationality. In both of these views there is a "higher life" of fulfilment contrasted by a "lower life" of sloth, irrationality, slavery, alienation, etc.[121] In each of these examples, it is the framework that allows human beings to distinguish between the "high" and "low" forms of life, and is what allows us to establish how far away we are from the "good life."[122]

While we have demonstrated that it is impossible to make judgements of worth without a certain framework (since choice then becomes arbitrary), what has not been shown is that considerations of worth are *necessary* for understanding human agency. Our example of the man fighting obesity gave us a reason to think that this might be true, but does not establish that this is necessarily the case. In fact, Taylor believes that the naturalistic tendencies found within much of our contemporary philosophy are bent on pursuing this very alternative: to give an account of our moral reactions without affirming any ontology of the human, that is, without acknowledging any moral framework.

Naturalism is the view that everything that exists is natural in kind, and that the acceptable methods of justification and explanation are commensurable with those found in science. In a naturalistic account of agency, moral frameworks are replaced by sociobiological or evolutionary explanations (like Dennett's), which attempt to show how or why moral reactions are possible.[123] Thus, if Taylor is correct, any naturalistic account of our discriminations between "right and wrong," "better and worse," "higher and

---

[120] *ibid.*, 23.
[121] *ibid.*
[122] *ibid.*, 42-45.
[123] Taylor, *Sources of the Self*, 5.

lower," etc., will be an inadequate starting point for understanding the self because no attention will be given to horizons of significance.[124]

So what we must pursue is the question of whether it is possible to understand human agency and the self without any consideration of an orientation towards a framework of *worth*. For the remainder of this chapter, I hope to make clear Taylor's claim that we cannot ignore our moral frameworks if we are to understand the self in any meaningful way. This will be accomplished by first examining the connection between strong evaluations and personal identity, and then by showing how such an understanding of personal identity presupposes an understanding of moral frameworks. In the fourth chapter I will outline the properties of the self that Taylor's methodology entails, and show why it is superior to Dennett's conception.

3.3 Identity and Responsibility

According to Taylor, responsibility is bound up with second-order desires because agents are responsible not only for the actions they perform, but also the evaluations they engage in.[125] We have certain *de facto* first-order desires, for example, the desire for ice cream, to play golf, to go to the cottage, for sex, etc, but second-order desires are not *given* in the same way. If a choice is made after the evaluation of certain first-order desires, then we endorse that choice because it is a decision that *we* have made based upon *our* considerations. We have an evaluated *motivation* or *reason* for making the choice we did. In this sense, second-order desires engage our responsibility not only

---

[124] The naturalist's motivation for ignoring the role of frameworks is based, among other things, upon the belief that the while we do have moral reactions, the postulation of moral *ontologies* is nonsense from a bygone age. Thus, the focus need only be towards moral *reactions*.
[125] Taylor, "What is Human Agency?", 28.

because our deliberations *cause* us to *act* in one way or another, but also because the *evaluation itself* is viewed as an activity.

To understand the strength of Taylor's claim about responsibility, we need to discriminate between taking responsibility for our weak evaluations and taking responsibility for our strong evaluations. This difference can be observed best by examining Taylor's discussion of Jean-Paul Sartre's notion of radical choice.

According to Taylor, Sartre claims that our values are created by the important choices we make. Consider the moral dilemma Sartre illustrates: a young man is torn between remaining with his ailing mother and going off to join the Resistance. For Sartre, there is no way to adjudicate between these two, so in the end the young man must make a radical choice; he must simply choose either one or the other. Sartre believes that whatever the young man chooses will determine what has value for him. That is to say, the choice the young man makes will dictate to the world who he is, because his action prescribes a certain value. The choice is valuable because he chooses it, nothing more.[126]

But Taylor's claim is that what is valuable for the young man is *already* given. The fact that the young man is in a *moral* dilemma in the first place tells us that he finds value in both choices. If Sartre were correct that our values are created by radical choice, then there could be no such thing as a moral dilemma, since a dilemma presupposes a difficult choice between two actions of antecedent value.[127] It would also be possible to have a "moral dilemma," according to radical choice, between staying at home with your ailing mother, and eating ice cream. Since value is created *after* choice, there is no *deeper* motivation to tell us that we *ought* to go one way or the other. Taylor would say that the

---

[126] *ibid.*, 29-30.
[127] *ibid.*, 30.

only type of responsibility enacted in Sartre's theory is that of a weak weigher. One is responsible for the choice made, but there is no responsibility associated with any kind of deep, articulable evaluation.

The deeper type of responsibility that Taylor wants to acknowledge can only be construed by the question of what one *ought* to do, conceived as something independently of our contingent preferences and desires. What adds this *ought* is the acknowledgement of a value inherent in the motivation for the choices within the dilemma; we are pulled in both directions because of the value each choice embodies. What one *ought* to do cannot be found in the method of radical choice since the choice is between two moral claims, and therefore, they necessarily contain a conception of worth or value *before* a decision is made. Since we are talking about *worthy* choices, then what is being considered falls into the category of strong evaluation, and with it, the kind of depth left out by the weak weigher in radical choice. With radical choice, value comes with the choice; the choice is valuable because you choose it, but the motivation for the choice is arbitrary. With deeper issues, "their importance is given,"[128] and therefore the motivation for choice lying behind our *moral* dilemmas contains a notion of quality.

Taylor argues that the type of responsibility associated with strong evaluations is intricately linked to personal identity.[129] It is in this sense that the evaluation of first-order desires takes us to the centre of who we are as agents. In fact, Taylor believes that without strong evaluations we would suffer a kind of identity crisis:

> Shorn of [certain evaluations] we would cease to be ourselves, by which we do not mean trivially that we would be different in the sense of having some properties other than those we now have ... but that shorn of these we would lose the very possibility of being an agent who evaluates; that

[128] *ibid.*, 33.
[129] By using the term "identity", I am referring to *who* we are as persons, our *selves*.

> our existence as persons, and hence our ability to adhere as persons to certain evaluations, would be impossible outside the horizon of these essential evaluations, that we would break down as persons, be incapable of being persons in the full sense.[130]

For Taylor, our identities are located within a sphere of questions that can only be addressed in strong evaluations. These "deep" questions define who we are, because they orient us within a framework from which we can determine what is good, valuable, meaningful, etc:

> To know who I am is a species of knowing where I stand. My identity is defined by the commitments and identifications which provide the frame or horizon within which I can try to determine from case to case what is good, or valuable, or what ought to be done, or what I endorse or oppose. In other words, it is the horizon within which I am capable of taking a stand.[131]

Without an orientation towards or within a moral framework, we lose the ability to situate ourselves within the frame of questions pertaining to worth, and therefore, we lose the ability to make strong evaluations. According to Taylor, if we are unable to say where we stand, then we lose our entire ability to evaluate, which not only omits our *depth* as selves, but also what makes us uniquely human. More importantly, we lose the ability to define ourselves morally as certain kinds of persons. For example, unless we countenance the idea of strong evaluations, defining oneself as a "homosexual" will have the same significance as defining oneself as "having brown hair" or being "five feet, ten inches tall." Taylor's claim, therefore, is that who we are (our identity) and what matters to us become *insignificant* without moral frameworks.

3.4 Self-Making and Responsibility

The discussion above focussed on the idea of *taking* responsibility for our actions as

---

[130] Taylor, "What Is Human Agency," 34-35.

well as the deliberations that lead to them. In light of Taylor's treatment of identity, however, it seems that, if who we are is constituted by "knowing where we stand in relation to certain types of questions," then by addressing these questions, we deploy a language of worth that makes our identities articulable. In this sense, the language we use is responsible for the *creation* of who we are at the deepest level.

Our motivations (our desires, aspirations, and evaluations) are formulated in words; we only have access to them as interpretations of feelings or behaviour.[132] Taylor believes that the way we choose to describe ourselves has an effect on the type of experiences we have as persons.[133] Recall Taylor's example of the compulsive overeater, and the two self-interpretations that were used to describe the motivation for avoiding the rich desserts. Articulating the avoidance of rich desserts as a desire to live longer, to have a good golf swing, or to get a date, creates an experience of the struggle against rich desserts that is significantly different from the one that articulates the avoidance of rich desserts as a matter of dignity. While the goal in both instances is to avoid rich desserts, the *motivation* is cast as a set of different desires, due to different interpretations of the situation. Taylor's claim is that if one re-describes one's desires (say, from the desire for dignity to the desire to maintain a smooth golf swing), then the entire *experience* of a situation is transformed because a new interpretation has been made.[134] For Taylor "our self-interpretations are partly constitutive of our experience," and therefore, "certain modes of experience are not possible without certain self-descriptions."[135] How exactly

---

[131] Taylor, *Sources of The Self*, 27.
[132] Taylor, "What is Human Agency?", 36, 36n.
[133] The term "description" is not being used in the traditional sense, i.e. as the neutral portrayal of a fully independent object. The description is an articulation; an attempt to make clear what is initially unformulated.
[134] Taylor, "What is Human Agency?", 36.
[135] *ibid.*, 37.

is this possible? And what are the consequences of this for our understanding of identity and responsibility?

According to Taylor, it is possible to make a change in both our experiences and our articulations. First, we can change our experience of a certain predicament through reflection, that is, by interpreting our experience in a new light. Changing the self-interpretation of my motivation for avoiding rich desserts from a desire for dignity to a desire to maintain a smooth golf swing will change my experience of the struggle from one that was articulable and *deep* (and potentially shameful) to one that is pragmatically inconvenient. Second, the richness of an experience may call us to change a description of it, because as it stands, the language may do an injustice to the depth or intricacies of the experience. For example, "to someone who strongly experiences the fight against obesity in terms of degradation, the 'deflated' descriptions appear a wicked travesty..."[136].

Therefore, in the relation between description and experience, a causal influence can run in either direction.[137] A change in the language of self-interpretation can lead to a change of experience, and the experience of a situation can lead to a change of description. Thus, a description of our motivations (our self-interpretations), and our experiences of what we hold important (our moral reality) are dependent upon one another. Because of this, Taylor argues that any attempt to articulate the motivations behind either of these (i.e. our language of worth, our moral actions or reactions), will be difficult, since the "object" we are attempting to be clear about does not "exist" independently:

---

[136] *ibid.*
[137] *ibid.*

> Our attempts to formulate what we hold important must, like descriptions, strive to be faithful to something. But what they strive to be faithful to is not an independent object with a fixed degree and manner of evidence, but rather a largely inarticulate sense of what is of decisive importance. An articulation of this 'object' tends to make it different from something than it was before.[138]

The articulation of what we hold important (our moral framework) is difficult because of the constitutive relationship of self-interpretation and experience. This relationship also ensures that what we hold important can never be exactly "faithful to reality." That is, the problem of articulation leads to the fact that some self-interpretations will be "more or less adequate, more or less truthful, more self-clairvoyant or self deluding,"[139] but they will *always* be interpretations nonetheless. But the fact that our interpretations will always distort the reality of the situation somewhat is what allows for the possibility of shaping the "object" (our moral frameworks) to make them accessible/inaccessible in new ways.[140]

Taylor suggests that the constitutive relationship between our self-interpretations and our moral reality engage our responsibility in two ways. First, we are morally responsible for what we come to see or fail to see as important after the interpretation of our experiences. That is, we are responsible for what we judge to be of importance in our lives. In the example of the man fighting obesity, he is responsible for coming to see either his dignity or the improvement of his golf swing as important. Secondly, and more fundamentally, we are expected to change ourselves for the better by attaining fresh insight into ourselves. Due to the depth considered in strong evaluations, and given the imperfections of personal character, there is always room for the re-evaluation of our

---

[138] *ibid.*, 38.
[139] *ibid.*
[140] *ibid.*

self-interpretations because there is always the possibility of error or delusion. Recall that the language we use shapes our interpretations of the world. Therefore, if we want our experiences to be more or less "faithful to reality," in the sense that we do not want to live a life that is delusional, we need to continually question our self-interpretations. We are responsible for the continual re-evaluation of what we consider to be important.

According to Taylor, our deepest evaluations, the ones that take us to the heart of who we are, are the "least clear, least articulated, most easily subject to illusion and distortion."[141] The type of evaluation Taylor is advocating is therefore radical, since one is questioning what is of fundamental importance to oneself. More importantly, what is being radically re-evaluated are the *terms* used in our fundamental evaluations (our notions of the "higher" life, what is a "worthy" cause, etc.), since these are used to articulate our motivations:

> ...in radical re-evaluations by definition the most basic terms, those in which other evaluations are carried on, are precisely what is in question. It is just because all formulations are potentially under suspicion of distorting their objects that we have to see them all as revisable, that we are forced back, as it were, to the inarticulate limit from which they originate.[142]

What is being questioned in a radical re-evaluation are the terms used to describe our motivations and experiences in a strong evaluation, since the language we use has a distorting effect on our motivations and experiences. We engage in a radical re-evaluation by using our familiar language, but we become more aware of how we are using it; we question the use of a familiar yardstick:

> ...the re-evaluation is carried on in the formulae available but with a stance of attention, as it were, to what these formulae are meant to articulate and with a readiness to receive any gestalt shift in our view of

---

[141] Taylor, "What Is Human Agency," 40.
[142] *ibid.*

> the situation, any quite innovative set of categories in which to see our predicament, that might come our way in inspiration.[143]

This type of evaluation engages the whole self, since these evaluations contain the language of worth in terms of which the self is organized. This is what makes the engagement particularly difficult.

By questioning the very terms on which our self-identity rests, we adopt a sort of "honesty policy" with ourselves. Only by questioning who we are at the deepest level can we be sure that our identities and what we hold important are as *authentic*[144] as possible, and at the same time reach deeper levels within ourselves. According to Taylor, it is only by going beyond the fixed yardstick by which we measure ourselves that we approach our full human potential:

> This radical evaluation is a deep reflection, and a self-reflection in a special sense: it is a reflection about the self, its most fundamental issues, and a reflection which engages the self most wholly and deeply. Because it engages the whole self without a fixed yardstick it can be called a personal reflection ... and what emerges from it is a self-resolution in a strong sense, for in this reflection the self is in question; what is at stake is the definition of those inchoate evaluations which are sensed to be essential to our identity.[145]

Thus, because our identities and experiences are partially constituted by our interpretations, Taylor believes that human agents are self-interpreting animals, since we must continually re-evaluate our interpretations to understand who we are (our moral frameworks). I will now show why Taylor's methodological approach to the self is superior to Dennett's.

---

[143] *ibid.*

[144] Taylor's concept of "authenticity" that will be discussed at length below. The meaning of this concept is implied with the use here, but the dictionary meaning of "authentic" can also work as well.

[145] *ibid.*, 42.

Chapter 4

Dialogism and the Limits of Naturalism: Re-evaluating the First-Person Perspective

On Taylor's view, any account of agency must countenance what we might call "our personal relationship with the world." Moreover, the starting point of Taylor's account of the self is the conceptual scheme of self-interpretation and self-understanding. The presuppositions of our self-understandings force us, Taylor thinks, to recognize that we are beings whose experience is structured by ineliminable moral frameworks. This dimension of the self simply cannot be ignored. Dennett on the contrary, being a naturalist, begins with the perspective of science, which forces him to consider the human agent as natural in kind just like any other object of scientific study. This does not entail, as we have seen, that Dennett is blind to moral issues; but they are not and cannot be central to his account. On this view, the moral dimension of our lives is something we must somehow accommodate within the scientific view. What needs to be further examined however is why this scientific perspective is an inadequate starting point for examining the self.

At the end of Chapter Two we noted two objections to Dennett's account. The first claims that because Dennett oversimplifies the relationship between the agent and interpreter, we cannot even recognize ourselves in his account. Understanding the self as an abstraction forces Dennett to omit the importance of our experience as agents, namely, the intimate relationship we have with ourselves, and the fact that we are agents with concerns. The second objection claims that the ethical dimension associated with Dennett's account is deflationary to a fault since it cannot properly explain instances of

incontinence. While these two objections do not constitute a knockdown argument against Dennett's account of the self, they do point towards an overall difficulty with Dennett's methodological approach. This difficulty has been accentuated in our discussion of Taylor, and we are now in a position to explicate it. In this chapter, Dennett's methodological approach to the self will be criticized for two specific reasons: i) his use of naturalism as a starting point fails when the object of study is the self, and ii) he cannot properly accommodate the dialogical nature of the self. I will discuss each of these in turn using Taylor's methodology in contrast.

4.1 A Problem for Naturalism

Taylor invokes the distinction between primary and secondary qualities, which has its origin in the seventeenth century, to describe the naturalist approach to the self. Primary qualities – such as shape, mass, extension, etc. – were considered to be the intrinsic features of objects. Secondary qualities were either thought to be sensory qualities not objectively belonging to objects but only seeming to belong to them, or properties that the object can be said to have only in relation to observers (i.e. colour). For example, without suitable perceivers to have certain experiences, there would be no such thing as colour. Therefore, subjective properties such as colour are suspect, according to the naturalist way of thinking, because they have no claim to be "real" constituents of the external world. A naturalistic approach to the self explains human action and agency in the same way it explains everything else. The focus on the self is made in physical terms, in a way that is completely independent of human experience.[146]

---

[146] Taylor, "Self-Interpreting Animals," *Human Agency and Language*, 46-47.

Taylor's methodology for understanding the self is radically at odds with this naturalistic method. Since personal identities are constituted in part by their self-interpretations, the self cannot be studied objectively since it is our interpretation of our own experiences and the terms we use to describe those experiences that determine who one is. But Taylor's methodology is not completely subjective either, for we saw that any interpretation made at the deep level of strong evaluation makes use of external moral frameworks, conceived as ontologies, and is therefore partially objective. How exactly does this relationship between the subjective and objective work, and what are the consequences of this for naturalism?

Recall that the constitutive relationship of our self-descriptions and our experiences make the articulation of our motivations (desires, aspirations, and evaluations) difficult, because the relationship between the two creates a distorting effect on the moral frameworks behind our motivations. However, according to Taylor, we have access to our deepest motivations through our emotions. For example, consider once again our moral dilemma: being faced with the choice of staying with your ailing mother or going off to join the Resistance. For Taylor, understanding much of the experienced motivation for either choice entails an understanding of the emotions involved.

The experience of an emotion can eventually lead to an articulation of our motivations, because emotions engage us with a "sense of our situation."[147] Experiencing a moral dilemma, for example, involves an experience of a situation with a specific property. This is not a neutral property, otherwise one could not be moved emotionally by the situation. Taylor calls this non-neutral property an "import:" the import serves as

---

[147] *ibid.*, 48.

the ground for emotion.[148] Thus, saying what an emotion is like involves making a specific import-ascription, a judgement that is either affirmed or denied.

According to Taylor, this analysis does not sit well with the naturalist picture and its notion of objectivity. Naturalists criticize the articulation of "our sense of the situation" for being analogous to articulations of secondary qualities: we are not describing "reality." However, Taylor believes that a completely objective approach will do violence to what is actually going on when we have emotional responses to certain situations. For instance, a naturalist approach may attempt to give a physiological description of an emotional experience, articulating it as a process of inputs and outputs. According to Taylor however, such an approach will only be properly able to account for import-less experiences. For example, a third-person behaviourist explanation of the experience of physical pain will work quite well, since the phenomena can be explained entirely by reference to some type of input/output model. Taylor's point is that we cannot make sense of *all* our experiences in such a manner. Since some experiences involve imports, then to even have those particular experiences must have required an interpretation of a situation. Thus, some experiences can only be explained by making reference to a world as interpreted by the subject.

For example, it seems that while one can objectify the import pertaining to fear, this is not possible for the emotional experience of shame. According to Taylor, a subject experiences shame only in relation to that subject's experience *as a subject*. Not only is this culturally mediated, more importantly, the experience of shame can only be had in relation to a motivation for dignity, pride, and respect:

---

[148] *ibid.*, 49.

> ...I may be ashamed of my shrill voice, or my effeminate hands. But of course it only makes sense to see these as objects of shame if they have for me or my culture an expressive dimension: a shrill voice is (to me, to my culture) something unmanly, betokens hysteria, not something solid, strong, macho, self-contained. It does not radiate a sense of strength, capacity, superiority. Effeminate hands are – effeminate. Both voice and hands clash with what I aspire to be, feel that my dignity demands that I be, as a person, a presence among others.[149]

According to Taylor, the linguistic meaning of "shame" can only be understood along with the emotional meaning of shame.[150] The emotion of shame can only be understood in relation to an appropriate import – something that warrants or calls forth the emotion of shame. Thus, the experience of shame can only be had in relation to concepts like dignity, pride, and respect. That is, a subject must interpret their situation in relation to these concepts in order feel shame.

For any naturalist conception of shame to be plausible, it must be capable of objectifying dignity, pride, and respect, as well as shame, but how can anyone objectify any of these? There is no such *thing* as "shameful," "dignified," or "proud" in the physical world, since these involve emotional meanings that surround our sense of certain situations. Without human beings, shame would have no meaning, and thus, this family of terms cannot be a natural, objective property of the world.

Any naturalistic approach to the self can only account for human experiences that do not involve the consideration of imports. But for Taylor, sense can only be made of terms like "shame" with reference to beings with purposes. But this manner of being is bound up with non-objectifiable terms such as "dignity" and "self-worth," as well as how others perceive us. By leaving out imports we leave out what is most fundamental to agency: what matters to us.

---

[149] *ibid.*, 53.

To see the strength of this point, we can ask if Dennett's theory can accommodate the meaning of emotional terms such as shame. At first glance, it seems that heterophenomenology can accommodate such terms. Recall, for example, Dennett's prediction that the lemonade seller will "exhibit surprise, blush, smite his forehead, apologize, and give me two cents" once I point out to him that he has given me the wrong amount of change. Now while this prediction entails the postulation of various beliefs and desires, heterophenomenology allows us to determine the narrow psychological states of the boy as well. Perhaps the boy tells us that he blushed and hit his forehead because he was *embarrassed* that someone his age made such a mistake. In our notional world we construct notional objects like dignity, respect, and self-worth, and explain the phenomenological experience of embarrassment by the boy's belief in this notional world.

However, if we are to appreciate the subject's emotional response we need to see it as appropriate to – as called forth by – the *embarrassment* of the situation. This requires that we think of the subject's response as more than simply a *reaction*. If Taylor is correct, then what is being articulated when the lemonade seller describes his situation as "embarrassing" is the boy's *interpretation* of the situation. As such, Taylor would argue that we must also think of the world that this emotion is a response to, as more than merely "notional."

Feeling shame or embarrassment is different from feeling pain, because the former are feelings that make reference to a framework of the good (what is dignified, what warrants respect, etc.) that is completely independent of the subject. This does not mean however, that a moral framework can then be "naturalized" to be thought of as an object

---

[150] *ibid.*

that can cause shame and embarrassment, in the same way a hammer can cause pain. This follows simply from a strict third-person methodology: our value terms represent a moral universe that is quite different from the one physical science claims to reveal and explain.[151] If we focussed on the *human universe* of what it is like to live as a human being we would lose the neutrality that is at the heart of naturalism.

Thus, according to Taylor, terms such as "shame" possess a "subject-referring" property, which is experience dependent.[152] It is Taylor's claim that since some of our emotions or experienced motivations involve import ascriptions, and since some of these imports are subject-referring, this offers insight into what the life of a human subject amounts to in a way that cannot be captured by the naturalist. Let us now examine further the connection between subject-referring emotions and strong evaluations, identity and responsibility, and self-interpretation. This will bring us closer to understanding what Taylor believes to be important for understanding the self.

As we have seen, according to Taylor the articulation of our situation as shameful, for example, is an interpretation. In fact, Taylor claims that our interpretation of the situation grounds the feeling of shame: "the feeling is what it is in virtue of the sense of the situation it incorporates."[153] While this "presuppose[s] a certain level of articulacy,"[154] Taylor believes that further clarifying articulations can always take place. Yet, if we clarify these feelings with the use of other terms, we can alter the way we feel. This follows simply from the constitutive relationship between self-description and experience. For example, our feeling of remorse "may dissipate altogether, if we come to

---

[151] Taylor, *Sources of the Self*, 59.
[152] Taylor, "Self Interpreting Animals," 54.
[153] *ibid.*, 63.
[154] *ibid.*

see that our sense of wrong-doing is unfounded."[155] The further articulation of our feelings opens up the imports behind them, and facilitates self-understanding.

This understanding of emotion ties everything we have been discussing together. Consider the connection between our emotions, our moral framework, and our personal identities. An articulation of an emotion with a subject-referring import will yield an expression of what is of importance, what matters to us, since what is really being articulated is a moral framework, what it is that gives rise to the emotional response of the situation in the first place. By articulating the framework, one is articulating one's "goods," what it is one finds worthy or unworthy. The articulation of one's moral compass gives an agent a chance to understand where she stands on certain issues, and therefore, who she is, what her identity is. But this also entails that moral frameworks, once articulated, can always be criticized and re-evaluated in light of new information, allowing for the possibility of changing how one feels, and what one finds important.

Since the naturalist cannot properly account for emotions like shame by objectifying either the agent or moral frameworks, the only recourse is to argue that we can understand the self without the consideration of any moral framework. The naturalist may argue that conceptions of "worth" or the "good" are human constructions that have been projected upon a morally neutral universe, and therefore a proper understanding needs no recourse to such conceptions. Like secondary qualities, we need not articulate horizons of significance in order to get to the *essence* of things. Recall however, that without any moral framework, we would be unable to make sense of terms like "dignity," "shame," "respect," "pride," etc. Perhaps then, the naturalist can argue that one does not

---

[155] *ibid.*

need these terms to understand the self. But as Taylor asks, what do we gain by being able to explain human behaviour without terms like "dignity?"

> Suppose I can convince myself that I can explain people's behaviour as an observer without using a term like 'dignity'. What does this prove if I can't do without it as a term in my deliberations about what to do, how to behave, how to treat people, my questions about whom I admire, with whom I feel affinity, and the like?
>
> But what does it mean 'not to be able' to do without a term in, say, my deliberations about what to do? I mean that this term is indispensable to (what now appears to me to be) the clearest, most insightful statement of the issues before me. If I were denied this term, I wouldn't be able to deliberate as effectively, to focus the issue properly – as, indeed, I may feel (and we frequently do) that I was less capable of doing in the past, before I acquired this term.[156]

If we ignore the terms by means of which we actually live our lives, then Taylor argues that we are ignoring significant aspects of our identities, since the terms we use help articulate who we are, what we feel to be important, etc. But then, are we not ignoring what it is to be a human agent? Are these not essential aspects of the self that we must attempt to explain? The naturalist distinction between appearance and reality fails when the object of study is the self, since what the self is, is a product of its own self-interpretations. As Taylor remarks: "...what ought to trump the language in which I actually live my life?"[157]

4.2 The Dialogical Self

According to Taylor, the self is essentially *dialogical*. A *monological* self-understanding involves a self-identity that is understood in isolation from its relations to others. The outside world and other agents form the content of the monological subject's "internal" representations, but the subject's self-identity is defined independently of

---

[156] Taylor, *Sources of the Self*, 57.
[157] *ibid.*, 58.

these. "The subject is first of all an 'inner' space, a 'mind,' to use the old terminology, or a mechanism capable of processing representations."[158] An understanding of the self as *dialogical*, on the other hand, requires that we conceive of the self as embodied, and as essentially incorporating the perspectives of others.

To understand the requirement of embodiedness, Taylor believes that we first need to understand that human agents are beings that are engaged in practices. Thus, we can only have an adequate understanding of the self as an agent *in practice*:

> As I sit here and take in the scene before me, I see a complex structure. It is oriented vertically, some things are "up," others are "down"; and in depth, some are "near," others "far." Some objects "lie to hand," others are "out of reach"; some constitute "unsurmountable obstacles" to movement, others are "easily displaced."...
> ...To say that this world is essentially that of this agent is to say that the terms in which we describe this experience – say those in the quotes in the previous paragraph – make sense only against the background of this kind of embodiedment. To understand what it is to "lie to hand" you have to understand what it is to be an agent with particular bodily capabilities that humans have.[159]

On Taylor's view, not only must we understand an agent with respect to the terms in which he or she lives his or her life, but we must also understand the agent in the world in which he or she lives. What needs to be explained are people living their lives, and this cannot be accomplished by removing the subject from the body of practices his or her life surrounds.

This leads to the fact that the self cannot be thought of objectively, which is to say, the self cannot be understood as an object like any other. According to Taylor, the explanation of an object that exists "objectively" has at least four properties:

---

[158] Taylor, "The Dialogical Self," in D.R. Hiley, J.F. Bohman, R. Shusterman, ed., *The Interpretive Turn: Philosophy, Science, Culture*, (Ithica: Cornell University Press, 1991), 307.
[159] Taylor, "Lichtung or Lebensform: Parallels between Heidegger and Wittgenstein," Philosophical Arguments (Cambridge: Harvard University Press, 1995), 62.

> 1. The object of study is to be taken "absolutely", that is, not in its own meaning for us or any other subject, but as it is on its own ("objectively").
> 2. The object is what it is independent of any descriptions or interpretations offered of it by any subjects.
> 3. The object can in principle be captured in explicit description.
> 4. The object can in principle be described without reference to its surroundings.[160]

The self cannot be seen as such an object. First, on Taylor's view, the self exists only within a sphere of questions that matter to the agent. One's identity, understanding "where one stands" on certain moral issues, is what allows a self to be oriented towards the good, in order to make evaluations. Thus, since a self cannot be understood independently of what matters to it, it cannot be taken "absolutely." Second, as we have already noted, the self exists on Taylor's view as a product of its own self-interpretations, and therefore it cannot be considered independent of these. Third, while the self is a product of its own self-interpretations, these self-interpretations can never be final. According to Taylor, "articulation by its very nature can never be completed."[161] We can only clarify one language by the use of another, and this "new" language can then be further unpacked and so on. Finally, the self must be considered in the context of its surroundings. When the focus is upon engaged agency, a background, a "context of intelligibility of experience"[162] arises with it, and cannot be separated from it.

Understanding the self as dialogical also requires that we think of the self as incorporating the perspective of others. This follows simply from our ability to use language:

> We become full human agents, capable of understanding ourselves, and hence of defining an identity, through our acquisition of rich human languages of expression....

---

[160] Taylor, *Sources of the Self*, 33-34.
[161] *ibid.*, 34.
[162] Taylor, "Lichtung and Lebensform," 71.

> ...But we are inducted into these in exchange with others. No one acquires the languages needed for self-definition on their own. We are introduced to them through exchanges with others who matter to us...[163]

According to Taylor, language "is not just a set of signs which have meaning in virtue of referring to something, it is the necessary vehicle of a certain form of consciousness, which is characteristically human."[164] On Taylor's view, we can only become a self through language, since this is our only access to certain facets of our self-understanding. But the point can also be made if we consider language in a broad sense, as a term that applies to other forms of self-expression as well (i.e. the "languages" of love, of gesture, of art, etc.).[165] A single agent cannot perform these acts of self-expression, since their activity requires at least an implicit awareness of social expectations and practices:

> Sawing and dancing are paradigm cases of dialogical actions. But there is frequently a dialogical level to actions that are otherwise merely coordinated. A conversation is a good example. Conversations with some degree of ease and intimacy move beyond mere coordination and have a common rhythm. The interlocutor not only listens but participates with head nodding and "unh-hunh" and the like, and at a certain point the "semantic turn" passes over to the other by a common movement. The appropriate moment is felt by both partners together in virtue of the common rhythm.[166]

It seems that Dennett's understanding of the self cannot be criticized for omitting "the other" in coming to understand what it is to be a human agent. Recall that Dennett accommodates the social embedding of the self as a process in the development of the mind, not only with the development of language, but more importantly, with the adoption of cultural memes of which the mind is composed. On Dennett's view, the

---

[163] Taylor, *The Malaise of Modernity*, 33.
[164] Taylor, *Hegel*, (New York: Cambridge University Press, 1975), 19.
[165] Taylor, *The Malaise of Modernity*, 33.
[166] Taylor, "The Dialogical Self," 310.

evolution of consciousness relies on ideas that are passed on through the processes of social interaction.

However, it is clear from our discussion of Taylor that Dennett's idea of selves spinning themselves out of narrative is wholly inadequate. An articulation of our situation as "shameful" or "embarrassing" reveals an interpretation, constituted at least in part by an implicit understanding of embodied action and a conception of the other. Without an embodied self-understanding there is no background to which "shame" or "embarrassment" can be made intelligible. Without at least an implicit understanding of patterns of appropriate action – "the dignified," for example – no interpretation of a situation can be "shameful." This implies an understanding of the other in our self-interpretations as well. A single agent cannot establish the background required for moral self-understanding. An action can only be described as "shameful" if there is a shared understanding between agents as to the meaning of "dignified." The meanings of our moral terms are introduced to us through our exchanges with others.

Dennett's conception of the self as spun by narrative fails to capture the dialogical structure of such a narrative. As Taylor has shown with the example of our moral self-understanding, the self is a necessary presupposition for the possible interpretation of dialogical patterns. Therefore, we must conceive of the self as a *spinner* – an interpreter – our experience as a self cannot be something that is simply *spun*. This means that Dennett's conception of "the other" is too thin – moral self-understanding cannot be spun in the vacuum of cultural memes.

The difficulty Dennett runs into, it seems, is how to understand the dialogical character of the self from broadly within the perspective of science. Where the

methodology of science begins is where the problem lies. Understanding the self as dialogical requires a simultaneous examination of both the self and other. On Taylor's view, the self can only be understood through processes of social interaction, but this important aspect is lost in a strict third-person methodology. The methodology of science, and therefore also, the intentional stance, must make the study of others *primary*, and then later on, incorporate this other-understanding into our self-understanding. This forces Dennett to remove everything associated with the first-person perspective (i.e. consciousness, morality, etc.), and then bring them back in later once they have received articulation in the understanding of others. It is clear why this does not work: When Dennett attempts to bring self-understanding back into the picture, the self is lost, due to the fact that everything associated with the self has now been defined in terms of the other. If Taylor is correct, then we need to define consciousness, morality, etc. by referring to the subject *as a subject*, since it is the interpretation of our own experiences and the terms we use to describe those experiences that define the self in the first place. The fact that we cannot recognize ourselves in Dennett's account is the result of other-understanding prior to self-understanding. What Taylor has shown is that a proper understanding of action and agency requires a methodological approach that is truly dialogical in nature.

Conclusion

This thesis has been an examination of the work of Daniel Dennett and Charles Taylor, both of whom approach the explanation of action and agency from the interpretationalist perspective. While both philosophies incorporate a third-person perspective, the methodology adopted by each leads to a different understanding of the self. Dennett's third-person perspective is empirically driven, following the model of natural science. Taylor's third-person perspective is premised upon the processes of hermeneutic and ethical reflection. Dennett's methodology begins with the agreed-upon methods of description and analysis of empirical science, thought to be neutral and unbiased, and proceeds to collect "data" from which his conception of the self is derived. Taylor's methodology on the other hand, begins with the world of the human being, with all of its interpretations, and proceeds to articulate what is required for self-understanding in such a world.

We can see the superiority in Taylor's position, first of all, by showing how Taylor's account of the self speaks to the objections raised against Dennett in the second chapter. The easiest way to do this is to consider where Taylor believes we find self-understanding: in the struggle for authenticity.

The question of "Who am I?" shows concern for the type of person one is or wishes to become. According to Taylor's view, this question cannot be answered properly without understanding the underlying moral force that lies behind it. For Taylor, to ask the question "Who am I?" is also to ask "What is a better or higher mode of life?" where "better" or "higher" correspond to a standard of what we *ought* to desire, and not simply

what we happen to desire or need.[167] To ask such questions takes us to the centre of who we are as agents, according to Taylor, for it is the first step in the quest for an *authentic* self-identity.

Since one's identity is essentially "knowing where you stand," one is therefore responsible for that identity and according to Taylor, this requires a constant evaluation and re-evaluation of "where you stand" if you are to have an *authentic* self-identity. Authenticity is therefore an incentive for a higher ideal, where the process of thought is one of reflection and self-question, towards the development of a free and unique individuality.[168] This is a rich understanding of the self that is socially embedded. On Taylor's understanding, authenticity requires that we recognize our relationships with others, which is the source for understanding one's moral framework. Without a conception of other-understanding permeating a conception of self-understanding, the life of the subject is "narrow and flattened."[169]

Now recall the two main objections raised against Dennett in the first chapter. The first had to do with our inability to see ourselves in Dennett's account, while the second shows Dennett's inability to deal adequately with the ethical dimension of human life. Let us examine the first objection and show how Taylor's conception of authenticity solves this problem.

How can Dennett claim to have the desired explanation of agency yet leave out the agent? Surely there is *something* to interpret from the intentional stance, but what is left out is the narrative of our lives told and acted out by *us*. With the self as an abstraction,

---

[167] Taylor, *The Malaise of Modernity*, 16.
[168] *ibid.*, 16-29.
[169] *ibid.*, 14.

we cannot conceive of a human life in this fashion. But then what we are left with must fail to count as an explanation of an agent at all.

Taylor's notion of authenticity encapsulates the human condition. Here we have a story about *us*, since what is in consideration is what matters to us, what our goals and purposes are, who we are and who we want to be. "In order to have a sense of who we are, we have to have a notion of how we have become, and of where we are going."[170] Such a rich narrative account is impossible for Dennett, since incorporating what matters to an agent breaks the bond of neutral terms to which he subscribes.

With regard to the second objection, Taylor is able to accommodate the ethical dimension in our lives in a way that Dennett cannot. With authenticity, our lives are focussed on the ethical dimension. By articulating our moral goods we orient ourselves to where we stand on certain moral issues. For Taylor, knowing where you stand is the first step towards self-understanding; to know who you are is to know where you stand in ethical space.

Due to Dennett's insistence upon the third-person perspective of empirical science, the ethical dimension of human life is seen as an extension of our rational decision-making procedures that have been developed through the processes of evolution. Thus, the idea of moral frameworks is repudiated in favour of rationality. Any distinction between "higher and lower," "good or bad," for example, is supposed to follow from what it is rational, and therefore, to choose the "higher" or the "good" is simply to do what a skilled deliberator ought to do. But as we have seen, this cannot be the case. Not only does such a view fail with inter-subjective moral terms such as "shame" and

---

[170] Taylor, *Sources of the Self*, 47.

"embarrassment" (which make reference to a world of dignity and self-worth – impossible to articulate on a naturalist conception), but it was shown that weakness of the will requires that we make sense of human beings as necessarily *irrational* creatures.

We are now in a position to articulate why Taylor's methodology is superior to Dennett's. A proper understanding of human action, agency, and the self seems to require a wider conception of what is *real* than natural science allows for. This finds several articulations in Taylor.

First, there is the idea that moral frameworks operate as a necessary background for the use and understanding of many of our moral terms. This view makes frameworks pre-existing and independent of ourselves as individuals, yet they cannot be articulated on any naturalistic conception of the world. Now while moral frameworks are obviously not a part of the natural world studied by science, Taylor believes it is a leap to go from this fact to the one that states that moral frameworks are not real and objective:

> Our value terms purport to give us insight into what it is to live in the universe of the human being, and this is a quite different matter from that which physical science claims to reveal and explain. This reality is, of course, dependent on us, in the sense that a condition for its existence is our existence. But once granted that we exist, it is no more a subjective projection than what physics deals with.[171]

If frameworks can be articulated, if they can have an effect on who we are as persons, and if we cannot do without them in using our moral language, then how can we say that moral frameworks are not real?

This leads into a second point. Not only are our moral frameworks real, but the fact that we must refer to them in order to make sense of ourselves as agents, shows that we therefore cannot understand ourselves without the use of certain terms. But these terms

---

[171] *ibid.*, 59.

necessarily incorporate the human universe and the human perspective, not the physical universe and the naturalistic perspective. I believe what Taylor has shown is that we cannot make sense of the world with the human perspective taken out. While Taylor's methodology attempts to expand our understandings of the human being in order to properly accommodate the scientific perspective, Dennett's methodology attempts to fit the human perspective into science. However, by attempting to fit everything within the naturalistic perspective, we lose our ability to see ourselves in the account. Our conceptions of freedom, autonomy, and responsibility are required for our moral self-understanding, and therefore, the meanings of these terms must remain intact if we are to properly understand ourselves. But as we have seen, Dennett's methodology forces him to change our understanding of these terms. Thus, given a strict third-person account of the self, we will never be able to see ourselves in the account, since the terms required for self-understanding have been removed.

In this thesis I have explored how Dennett's naturalism forces him to find a way of accommodating the mind, self, value, etc., within nature as science understands it. On Dennett's view, it simply makes evolutionary sense that there should be beings who come to think of themselves as minded, conscious, selves. I conclude, however, that it is very difficult to see how Dennett can offer the kinds of explanations he does without seeming to explain away aspects of our lives that have an enduring presence, indeed authority, for us. While for Dennett science sets the explanatory standards, for Taylor the benchmark is always self-understanding. Taylor always starts by exploring the preconditions of the central ways in which we see ourselves. The best explanation of ourselves, according to Taylor, begins by explaining the contours of our conceptual scheme. So for him, there is

no question of trying to legitimate or accommodate those understandings in relation to science, for nothing science can do could undermine them. Taylor would argue that if science says something different from what follows from his methodology, so much the worse for science. Since our self-understandings are so basic to our identities, Taylor argues, anything that purports to conflict cannot carry real conviction. In this sense, it might be said, Taylor is more of an interpretivist than Dennett, for it is not just that our mental lives are, in some sense, constituted by our modes of interpretation. It is that all forms of understanding, even science, are, at least in part, exercises in self-understanding, and are all in that regard equally interpretative. It is, I conclude, Taylor's version of interpretivism, rather than Dennett's, that offers us the best chance of attaining a satisfying conception of human agency and of our place in nature.

Bibliography

Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell.

Bruner, J. (1995). "Meaning and Self in Cultural Perspective," in D. Bakhurst and C. Sypnowich, ed., *The Social Self*, London: SAGE Publications, pp. 18-29.

Davidson, D. (1980). "How is Weakness of the Will Possible?" *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 21-42.

Davidson, D. (1984). "Thought and Talk," *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 155-170.

Davidson, D. (1986). "A Coherence Theory of Truth and Knowledge," in E. Lepore, ed., *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell, pp. 307-319.

Dahlbom, B., ed. (1993). *Dennett And His Critics*, Oxford: Basil Blackwell.

Dennett, D.C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge: Bradford Books/The MIT Press.

Dennett, D.C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge: Bradford Books/The MIT Press.

Dennett, D.C. (1987). *The Intentional Stance*, Cambridge: Bradford Books/The MIT Press.

Dennett, D.C. (1991). *Consciousness Explained*, Boston: Little, Brown.

Dennett, D.C. (1996). *Kinds of Minds*, Great Britain: Weidenfeld & Nicolson.

Dennett, D.C. (1998). *Brainchildren*, Cambridge: Bradford Books/The MIT Press.

Hacking, I. (1995). "Why Multiple Personality Tells Us Nothing about the Self/Mind/ Person/Subject/Soul/Consciousness," in D. Bakhurst and C. Sypnowich, ed., *The Social Self*, London: SAGE Publications, pp. 159-179.

Knight, D. (1992). "Theory of Action and the Interpretation of Cinematic Narratives," *Stanford French Review*, 16.2, pp. 197-213.

Knight, D. (1994). "Selves, Interpreters, Narrators," *Philosophy and Literature*, 18, pp. 274-286.

Ricoeur, P. (1992). *Oneself as Another*, trans. K. Blamey, Chicago: The University of Chicago Press.

Searle, J. R. (1995). *The Construction of Social Reality*, New York: The Free Press.

Stich, S. (1981). "Dennett on Intentional Systems," *Philosophical Topics*, 12.1, pp. 39-61.

Stich, S. (1985). "Could Man be an Irrational Animal? Some notes on the Epistemology of Rationality," in H. Kornblith, ed., *Naturalizing Epistemology*, Cambridge: The MIT Press.

Taylor, C. (1964). *The Explanation of Behaviour*, London: Routledge & Kegan Paul Ltd.

Taylor, C. (1975). *Hegel*, New York: Cambridge University Press.

Taylor, C. (1985). "What is Human Agency?" *Human Agency and Language, Philosophical Papers Volume One*, Cambridge: Cambridge University Press, pp. 15-44.

Taylor, C. (1985). "Self-Interpreting Animals," *Human Agency and Language, Philosophical Papers Volume One*, Cambridge: Cambridge University Press, pp. 45-76.

Taylor, C. (1989). *Sources of the Self*, Cambridge: Cambridge University Press.

Taylor, C. (1991). *The Malaise of Modernity*, Concord: House of Anansi Press Limited.

Taylor, C. (1991). "The Dialogical Self," in D.R. Hiley, J.F. Bohman, R. Shusterman, ed., *The Interpretive Turn: Philosophy, Science, Culture*, Ithica: Cornell University Press, 304-314.

Taylor, C. (1995). *Philosophical Arguments*, Cambridge: Harvard University Press.

Williams, B. (1985). *Ethics and the Limits of Philosophy*, London: Fontana Paperbacks and William Collins.