


Spring 2015

## The Mind, the Brain, and the Self: The Limits of Sense and Nonsense in Neurology and Psychology

Max Boris Baird  
*Bard College*

Follow this and additional works at: [https://digitalcommons.bard.edu/senproj\\_s2015](https://digitalcommons.bard.edu/senproj_s2015)

 Part of the [Philosophy of Mind Commons](#), [Social Psychology Commons](#), and the [Theory and Philosophy Commons](#)



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).

---

### Recommended Citation

Baird, Max Boris, "The Mind, the Brain, and the Self: The Limits of Sense and Nonsense in Neurology and Psychology" (2015). *Senior Projects Spring 2015*. 116.  
[https://digitalcommons.bard.edu/senproj\\_s2015/116](https://digitalcommons.bard.edu/senproj_s2015/116)

This Open Access work is protected by copyright and/or related rights. It has been provided to you by Bard College's Stevenson Library with permission from the rights-holder(s). You are free to use this work in any way that is permitted by the copyright and related rights. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. For more information, please contact [digitalcommons@bard.edu](mailto:digitalcommons@bard.edu).

The Mind, the Brain, and the Self:  
The Limits of Sense and Nonsense in Psychology and Neurology

Senior Project submitted to  
The Division of Social Studies

by  
Max Baird

Annandale-on-Hudson, New York

May 2015

## Table of Contents

	Page Number
Chapter One	
Section A: Introduction	1
Section B: Paul Churchland's eliminative materialism	5
Section C: Thomas Nagel's neurological correlates	13
Section D: Peter Hacker's psychological concepts	21
Chapter Two	
Section A: Introduction	27
Section B: Why does language have any relevance to empirical work?	29
Section C: Wittgenstein's defense of criterial meaning	36
Section D: Three examples of Hacker's method in response to empirical studies	47
Section E: Are Libet's results impossible to interpret?	53
Chapter Three	
Section A: Introduction	57
Section B: Criterial meaning versus ostensive reference?	62
Section C: Prediction and Usefulness	69
Section D: Psychological concepts, post-folk psychological theory, and a third option	73
Section E: Applying our psychological topics of interest	78
Section F: Conclusion	83
Works Cited	89

Chapter One

## Section A: Introduction

Do you remember in 10th grade when Morgan broke up with you? You instantly started crying in the hallway. Her words took a moment to process, but when you jerked your head up to meet her eyes, the certainty therein hit you like a cafeteria tray full of hash browns in a food fight. It was a sort of disembodied feeling, as though, no, this could not really be happening. And yet when you looked in her eyes there was no doubt in your mind that it was. Just as the bell rang and students began streaming out into the hallway, you burst into tears. You knew why you were crying - Morgan was the love of your life, and without her you would never be complete. Life could not go on.

And yet imagine if a stranger in a lab coat had been strolling down the hallway, and, at the moment you comprehended what Morgan had communicated and the waterworks began, this stranger told you, “While it may seem that you are in a state of heartbreak in response to what Morgan expressed, this claim requires an incorrect assumption. You do not have the grounds to claim that your crying or your emotional reaction is in response to your break up. What causes emotions, much less what causes physiological reactions like crying, is not a determination that is available to introspection.” Then or now, this statement might strike you not just as rude, but also as truly bizarre. The moment you processed the words you felt the feeling of heartbreak and you began to cry. Even if, in the heat of your youth, you overreacted, that makes the statement seem no less impossible. Could there have been some other causal factor that you were not aware of that triggered you to start crying, which just happened to occur at

the same time? Was your feeling of sadness, shock, or betrayal, that itself seems to be such a specific response to such a specific stimulus, actually illusory?

If it were contemporary psychologists Nisbett and Wilson in lab coats walking down the hallway, they would urge you to distrust your judgments about what causes you to you feel, think, or act. They believe that we are frequently unaware of what stimuli affect us, when they do so, or that they have done so. Participants in studies on motivation are unaware of how stimuli affect their motivation or that a certain stimuli had affected them (Nisbett and Wilson 237). People are unaware of how associations with certain words influence their responses (the phrase “ocean-moon” primes participants to prefer Tide brand detergent), or how the order of stimuli affects participant’s responses (the item most to the right in a shop window is preferred by a factor of four to one) (240). When participants try to report how stimuli affect them, their answers are so far removed from the processes that investigators have rigged the experiment to trigger that participants seem to have no direct access to that process (238). Moreover, the answers we do give via introspection are misguided in the same predictable ways, such that introspection does not seem to be personalized at all. When we explain ourselves, we depend upon what Nisbett and Wilson call “*a priori* causal theories” (248). We report stimuli that seem plausible to explain our behavior according to what our culture, sub-culture, or individual network of associations has implicitly or explicitly suggested to us is plausible. The strength of one phenomenon’s variation with a mental variable (such as whether Morgan wants or does not want to break up and you feeling sad or not sad) gives us no clue as to whether or not we will be aware of that covariation. If a cultural theory correctly describes a relationship between two variables, then we are lucky.

One such relationship is that break ups cause heartaches. Is it a matter of chance that we say break ups cause heartbreak? The definition of heartbreak is that it is pain from romantic difficulties – how could it be a matter of chance? What would heartache be if it did not follow from break ups? Maybe what we call heartache is no different from other sorts of emotional pain, and what causes heartache is no different from the cause of many other emotional states. Of course, when Morgan spoke those fatal words to you, you did not feel nothing. What you felt was a very specific feeling, right? (Did it not feel that way?) But if we are willing to say that the cause of heartache can be the same as the cause of a number of other mental states, then we may have to change an important part of the way we define heartache.

Although Nisbett and Wilson are careful to note aspects of our introspection that their argument does not apply to, such as our ability to determine what we are currently feeling, what we are sensing, and what we are attending to (255), their argument should still be shocking. With regard to knowing psychological causes and effects - a fundamentally important determination - the way our minds work seems to be nothing like we imagined. And as more psychologists and neurologists explore how our minds function, the norm seems to be that these scientists are able to not just explain how we function, but are able to explain how we function better and differently from how we normally believe we function. This is not just a question of convenience, a matter of which type of explanation will get the best results. Empirical investigation may not just change how we relate to our emotions, our free will, our self-awareness, and all the topics that constitute our mental lives. Empirical investigation may cause us to completely abandon the topics that we are currently familiar with. It may describe how we function

in ways that will be totally alien to the psychological words and concepts that we currently possess.

In this essay, we will explore different possible relationships our psychological vocabulary can have to empirical research. Specifically, we will examine one philosopher, Paul Churchland, who believes that, by empirically investigating how the brain functions, we will be able predict and explain how we function better than our current psychological theories can. Our psychological concepts will not conform to how we neurologically function, and so they will eventually be eliminated. In contrast, Thomas Nagel believes that psychological concepts, insofar as they serve as the names of experiences must be empirically identified with neurological states. If we have an experience, there must be a neurological basis for that experience, and so if we define our psychological concepts in terms of what it is like to experience them, then these psychological concepts must have some neurological basis. Instead of believing empirical work will eliminate our psychological concepts, Nagel believes that, in principle, empirical work should validate our psychological concepts. The philosopher Peter Hacker, along with a neuroscientist M.R. Bennett, will present difficulties with both accounts; neither, they think, will be able to specify a satisfactory relationship between our full psychological vocabulary and how empirical research is and will be conducted. This first chapter, by exploring different conceptions of how we may be able to describe ourselves once we know more about how we function, will specify what is at stake in this debate and how Hacker's concern with language can figure so prominently in it.

## Section B: Paul Churchland's Eliminative Materialism

Paul Churchland argues that a change in the way we describe ourselves is not just possible but necessary. He presents a stronger, broader argument than Nisbett and Wilson do. Put bluntly, he says, "Our common sense conception of psychological phenomena constitutes a radically false theory" (Churchland 67). When we make an introspective judgment, report a sensation, or think a thought, we are attempting to describe some process, however complex and abstract, occurring in a brain. Our capacity to explain and predict our own and other people's behavior and mental states, our "folk psychology", does a poor job of this. By putting our folk psychology within propositional statements, we can see how folk psychology operates in a law-like way. A few examples of this are: 1) that if people suffer bodily damage, they generally feel pain, 2) people who are angry are generally impatient, 3) people who fear P generally hope that not-P, and 4) people who desire P and know that Q is a means to P will generally desire Q (61). These rules may be a handy guide to understanding how people will behave, but Churchland wants us to think of these rules as hypotheses about how we function. Do these rules do the best job of explaining and predicting how we behave or how our cognition occurs? A folk psychological rule about how we function is that people who experience break ups undergo heartbreak. We may have some evidence for this rule – we probably would not think it if it did not seem to explain how people tend to behave after break ups. But this is very different evidence than the evidence we would need to develop a neurological theory of how our brains function after break ups.

As a theory, folk psychology gains authority only by virtue of being the best hypothesis available to explain and predict human activity (69). However, there are



plenty of activities that human engage which folk psychology cannot explain or predict, as Nisbett and Wilson's paper repeatedly shows. These failures might not seem so egregious if folk psychology were adapting in response to them. Instead, Churchland sees folk psychology retreating. Where "primitive" cultures explained natural phenomena by using folk psychology (the wind could be angry and oracles could augur the future), we now have a variety of other sciences at our disposal that can do a better job of explaining and predicting these phenomena (74). Even when only applied to the "higher animals", folk psychology displays a remarkable independence from our other fields of scientific knowledge, particularly neuroscience (75). Folk psychology's explanatory gaps, stagnation, and independence from other sciences all suggest it may be replaced the same way chemistry replaced alchemy.

Once we have a better theory as to how we actually function – based on observations of how our brains function – we will develop a completely new language to describe our mental states and capabilities, one that does not at all resemble our current vocabulary of psychological concepts. The phenomena Nisbett and Wilson identify seems to be one area of our folk psychology that is deficient, but Churchland's argument goes further than noting deficiencies. Not only will our folk psychological generalizations change, such as our association between anger and impatience. Our psychological concepts themselves eliminated and replaced. All our ways of describing ourselves, from describing our emotions or personalities to our concept of free will, stake a claim as to how humans function (87). Whenever we use these terms in our daily lives, Churchland believes we are doing a poor job at theorizing about how we function. By providing an empirically validated description of ourselves, we will so successfully and so differently

describe ourselves that no one will use folk psychology or our psychological concepts anymore.

To begin to get a sense of what Churchland is conceiving, imagine that you were told that heartache is better understood as a loss of a sense of security. What sort of evidence might lead someone to make a claim like this? Perhaps the mechanism that produces either emotional state is the same, or perhaps we could predict the intensity or duration of someone's heartache based on the severity of their loss of security. This example is not so hard to swallow. Describing heartache as a loss of a sense of security might even make sense without empirical evidence - although there are situations where this might not apply, in others it might be an astute observation. On the other hand, imagine that someone told you that our concept of melancholy is misconceived, and that melancholy actually is impatience. This is much harder to rationalize with folk psychology. Despite that, in both cases we have only put forth concepts contained within folk psychology. Churchland's argument is not that we are just missing access to how our mental states are related to other variables, as Nisbett and Wilson argue. Even deeper, he believes that we do not even know the correct terms to refer to how we function. As we begin to use empirical evidence to describe how our brains and bodies function, we will cease to discuss categories of phenomena as basic as emotions, intentions, or personalities. This is what Churchland calls eliminative materialism or eliminativism: once we better understand how we function, our folk psychological theory of how we function will be displaced, not replaced, and ultimately eliminated. As Churchland says, the onus is not on "empirical systems to instantiate faithfully the organization that (folk

psychology) specifies” (78). Instead, the onus is on our psychological vocabulary to instantiate our empirical systems.

It does not matter to Churchland if this seems difficult to imagine. Our lack of creativity does not limit what we can discover empirically. He himself has a couple of suggestions for how communication could occur beyond folk psychology, although he is quick to emphasize that they are thoroughly speculative. If we are trying to describe an object, for instance, then we may be able to give a far more informative description of that object by describing the actual neurological process by which we do perceive that object. One proposition he suggests, then, is a new language based around these “internal structures” of our brains (87). The efficiency, quantity or accuracy of the information communicated may serve as a yardstick by which we can say that this new, empirically derived language is in fact a better theory than folk psychology. Churchland’s second proposition, even stranger, is that we could develop a means of communicating from one brain to another in the same way that the two hemispheres of our brain communicate (88). He outlines a technology that would allow us to communicate without that communication relying on folk psychology. He imagines an artificial commissure modeled after the corpus callosum, the band of nerves that connects the two hemispheres in one person’s brain, but made of microwaves transmitting information from a neural implant. If the technology is speculative, the improvement he describes here is very specific. We could empirically measure the difference in quantity of information transmitted. Information is sent between our two hemispheres at a rate of  $2 \times 10^8$  binary bits per second, whereas spoken English communicates less than 500 bits per second—a dramatic improvement.

A post-folk psychological theory will replace the psychological topics we currently communicate about because it will change the rules by which we believe we function. Will artificial commissures do that? When we talk about a connection between two brains, communication seems to be the wrong word to describe how information is exchanged between them. It might be more appropriate to say that this direct, instantaneous connection will make communication unnecessary. This is then a different way in which the replacement of folk psychological rules, as well as communication about psychological concepts, will be eliminated. In Churchland's second suggestion, communication is not so much replaced as it is made obsolete. When we say that our two hemispheres "communicate", we are really speaking about the activity that occurs in our corpus collosum, which is not exactly communication as we normally use the word. In the first case that Churchland outlines, in which we communicate by describing the activity that occurs in our brains when our brains perform a certain task, we still have to "figure out" what the person means, so to speak. If someone describes the process that occurs when they perceive the color of a physical object, than one would still have to know how to interpret the neurological processes that the person describes. Explaining how this sort of language might develop, Churchland writes,

(I)t is not inconceivable that some segment of the population, or all of it, should become intimately familiar with the vocabulary required to characterize our kinematical states, learn the laws governing their interactions and behavioral projections, acquire a facility in their first-person ascription, and displace the use of (folk psychology) altogether. (86)

Some theoretical background would be needed in order to understand the operation of "kinematical states". Whether this knowledge becomes commonplace is an open question, despite all the possible advantages of this new form of communication.

In contrast, no theoretical background is needed for the direct transmission of information bits from one brain to another. You do not need to know how your own brain's hemispheres communicate in order for them to do so, and you do not need to know how gravity works in order for objects to fall. Transmission of information via an artificial commissure likewise would not require any shared conventions as to how to interpret the information in order for the technology to work. This transmission is communication in the sense that information is being exchanged, but it is a peculiar sort of communication that requires nothing in order to be understood. This transmission just happens, regardless of the receptiveness of the "listener".

Will this change the rules by which we think we function, or in other words, how we describe ourselves? One might object that if we knew how to create the technology, we would have to have some theoretical background as to how that machine worked. Will this theory about how artificial commissures work become a theory about how we function? Can it provide a radically new way to describe what happened to you in that fateful, high school hallway? We have no assurance that, once we are able to instantaneously transmit information between each other, we will stop transmitting information about our current psychological topics. What Churchland still needs to posit is that whatever theory explains how this artificial commissure works is also a theory about how we, our brains and bodies, function. We are something like computer programs with a heuristic guide to how we function, on the verge of discovering our actual computer programming. However, we need to know that there is some new theory that can displace our old heuristic. Assurance that other concepts can replace folk psychology is vital for Churchland's argument—if no new theory could do for us what

folk psychology does, then folk psychology might not be a theory waiting improvement in the same way chemistry was an improvement in explanation over alchemy.

Consider the problem this way: will it be correct to call post-folk psychological theories psychological? When we use the neurological states and processes that allow us to perceive in order to describe objects more accurately, Churchland believes that we can use the mechanics of our brains to displace folk psychological explanations of how perception occurs. Although we might improve how we describe perception, it seems peculiar to say that we would eliminate talking about perception. This new way of describing perception would rely on identifying and describing the mechanisms that produce perception, but we would never stop describing perception. It is hard to imagine any post-folk psychological set of concepts that rejected and displaced our psychological concept of perception. Perhaps we need to draw a distinction between folk psychological rules and psychological concepts, such that certain our folk psychological rules will be replaced, while some psychological concepts will be neurologically described in a post-folk psychological language and others will be eliminated.

This would of course complicate Churchland's argument, and we will return to this idea in chapter three. For now, all we can say is that Churchland expects new post-folk psychological theories to be more useful to us, but he does not question whether these better fulfilled uses are folk psychological in origin. Because the improvements he imagines are limited to prediction and quality of explanation, he might not have to. Predicting one variable on the basis of another or providing a mechanical explanation of how a brain functions can be done without any reference to folk psychology. Regardless of the topic, whether we are discussing free will, perception, or heartbreak, a post-folk

psychological language can improve our current language in these two ways, by better predicting and explaining how we function.

Churchland's argument, then, may not just be about the types of concepts we use to communicate. The concepts we use will change because the reasons we have for using concepts will change: our theories about how we function will change from being relatively unfounded, useless conventions to being empirically validated and, as a result, dramatically more useful. We will see that Hacker believes we cannot characterize our current method of communicating as contingent upon unfounded, useless conventions. He will argue that prediction and explanation cannot do justice to the variety of uses he believes are present within folk psychology. For now, however, the force of Churchland's argument is unchanged. Nisbett and Wilson, in one paper, identify swaths of phenomena folk psychology cannot account for. What should stop us from finding another theory that could?

The prospect of doing so might put us ill at ease. When Morgan broke your heart, did you think you were heartbroken just because of your folk psychological belief that heartbreak follows break ups? That is an expectation we have; that is true. But of course you did not just think, know, or believe that you were heartbroken. You felt it too. Is it really because of a lack of creativity that you feel so certainly?

### Section C: Thomas Nagel's Neurological Correlates

Thomas Nagel is one philosopher who does not think so. In his article, "What Is It Like To Be A Bat?", he explores what would be required in order for our conscious awareness of our experiences to be identified with certain neurological states. Nagel's focus is not so much on empirically validating this awareness as it is on finding a

neurological correlate to our experiences, identifying a certain brain state that we can say is the brain state of a person undergoing a specific experience. If you experienced heartbreak, then how that feeling consciously felt must have some corresponding brain state. Even when we cannot articulately describe some experience, there is something that it is like to have experience that we privately undergo. Whatever this is like, these brain states must somehow be the feeling of undergoing experience X. We must somehow be able to say that “intrinsically, there is something that it is like to undergo a physical process” (Nagel 445). However, this does not lead Nagel to believe that our current descriptions of our experiences are wrong, as Churchland might anticipate. Instead, Nagel does not believe that we currently possess a straightforward theoretical background through which we can understand how an empirically validated state of neurological affairs is an experience. To say that an experience just *is* some neurological state would be to ignore that our experiences feel like something when we undergo them. Nagel compares this to how we might use the word “is” when we say that a caterpillar is a butterfly. We have to specify how we mean the word by describing the whole process of metamorphosis – without that explanation, it would not be informative to predicate a caterpillar with a butterfly. Likewise, if we ask, “What is experience X?”, and get the answer that it is some specific neurological state, we may still be left wondering what that experience is like for someone undergoing it. We need to find a productive way to identify experiences with neurological states (447).

Nagel provides us with a specific formulation for how to describe our experiences. We have an experience when there is something it is like to be or do X that you are aware of at that moment. This covers a lot of ground—we can say that there is



something it is like to smell, something it is like to have a cold, or something it is like to be human. Although these phrases' referents are to different degrees experiences that are felt at specific moments, if there is something that is like to undergo them then Nagel would consider them equally prone to physical identification. Nagel asks us to imagine what it is like to be a bat—this may be an impossible task to succeed at, but we nevertheless do believe that bats undergo experiences and are conscious of things. What is it like to have sonar? I have no idea, but I do believe that I know what it is like to see. If I have visual experience, and bats are no different than humans in their ability to be conscious of things, then why should bats not have sonar experience? Yet what “sonar experience” is may be impossible to comprehend without the sensory equipment that sonar requires (438). The difficulty in identifying experiences with neurological states is especially apparent here. If we do not know what sonar experience is like for bats, we will not know how to understand how any physical state could be sonar experience.

Our ability to explain what experiences are like is thus limited. Nagel believes there are certain experiences of which we do know what they are like, (experiences you have undergone), others we do not (bats' experience of sonar, for instance), and all sorts of experiences in between. These are experiences that you may not have undergone yourself, but which you are nevertheless able to understand in some sense by extrapolating from your own experiences (439). If I have not experienced heartbreak before, I may not know what your experience of it was like, but I could extrapolate from feelings I have had of sadness and loss in order to get some sense of it. Although we need to undergo experiences first hand to know what they are truly like, we can still access experiences we have not undergone indirectly, through extrapolation. We cannot, on the

other hand, extrapolate from our experiences to learn about a bat's experience of sonar. Our physical constitution, in this case our sensory equipment, is too different.

In cases in which we can use our own experience to extrapolate, we encounter degrees of success all the time. Describe your high school trauma to me, and I'll be empathetic. "That must have been awful," I say, and you agree; it was. So we seem to have found some point of agreement concerning that experience and identified some aspect of it. This is not just a correct application of folk psychological theory, unfounded by empirical validation. It is a statement about what the experience is like to undergo, and insofar as you felt a feeling of awfulness, than that awful feeling must have some physical basis. Unlike Churchland, who believes that the onus is on folk psychology to conform to what is science empirically validates, Nagel believes that the onus is on science to explain how our experiences can be physical states.

To the extent that our names of experiences are aspects of folk psychology, has Nagel found a category of phenomena that Churchland did not consider? Perhaps we could we say that the two arguments apply to different "sections" of our folk psychology. Nagel takes it for granted that our descriptions of our experiences accurately describe those experiences. After all, we are able to better describe our own experiences than we are able to describe other creatures', or even other people's, experiences – the trouble is how we can communicate the same information by discussing neurological correlates. Churchland agrees that we know folk psychological descriptions of our mental states and processes, but that these descriptions do not meaningfully refer to anything. So maybe there are some cases in which we will empirically verify how we describe ourselves, like whenever we describe our conscious experiences, and others where our current self-

descriptions will be replaced, like when we describe our mental processes in order to make some prediction about what we will do, feel or think.

However, consider how peculiar it would be to divide up a concept like decision-making into parts that Churchland could deal with and parts that Nagel could. Churchland could point to work like Nisbett and Wilson's, and describe decision-making as a folk psychological theory while providing better ways to describe it. At the same time, Nagel could say that we do have sudden moments of clarity in which we experience decisiveness. We experience weighing the pros and cons of a decision, and we experience indecisiveness. Is it possible that in Nagel's neurological correlates, we could find an empirical defense not just of consciousness, but of folk psychology?

This might be a bizarre way for science to direct itself. Churchland would quickly point out that what we are imagining is a science no longer based on explaining and predicting how we function. Maybe we could find a neurological correlate to heartache, even to your particular, individual experience of it. Maybe we could see to what extent that experience is composed of different types of pain, and maybe we could see to what extent your heartache is similar or dissimilar to other people's experiences. So there might be ways that we could creatively verify our folk psychological stories through identifying neurological correlates to our conscious experiences. But just because this is possible does not mean it is ideal or preferable as a research program. We may have neurological correlates to our experiences, but the experiences we are conscious of may still be misleading, as Nisbett and Wilson show. Churchland thinks that this is because we cannot make any predictions relating those experiences to anything meaningful within our folk psychology. If our experiences are to be significant, then we must be able to

generate explanations or predictions that will support how we currently believe we psychologically function. Consider promise making. It would seem a particular sort of promise that was made completely unemotionally, without a feeling of commitment. If a friend makes you a promise in a flippant way, then you will be less inclined to believe they will keep that promise. But then imagine we are told that this feeling of commitment actually tells us nothing about whether a person will or will not keep their commitment. We would no longer be interested in how earnestly a promise was made if we were trying to determine the likelihood of that promise being kept. In the same way, even if you are right in saying that on January 20<sup>th</sup>, 2010, at 11:45 AM between Trigonometry and English you felt heartbroken, the rest of the associations we have with the concept of heartbreak, such as whether it was caused by your break up, may be misleading. In effect, Churchland might respond to Nagel, fine, conscious experience might have some empirically verifiable neurological correlate, and so our descriptions of our experiences may be validated in some way. But if that is all we can say about our experiences, then it is not enough.

So neurological correlates may not provide the opening for us to make predictions that will support folk psychology. Still, other strategies may allow neurological correlates to validate folk psychology. If prediction is no good, we might still make headway by comparing similarities and dissimilarities between neurological correlates. After all, we have different experiences, and those differences ought to be somehow neurologically instantiated. Folk psychology might be misguided in terms of prediction, then, but it might be not fundamentally misguided in the way Churchland imagines. Indeed, Nagel's process of extrapolation, when it is applied to other humans, may be entwined with folk

psychology. If someone tells you that they were just dumped, you generally have some idea of how that person is doing and what they are feeling. As Nagel explains, you do not need to undergo another human's experience in order to have some sense of what that experience is like. When I say that I know how difficult your break up was despite not having gone through it myself, that process of extrapolation itself seems to involve an awareness of a relationship between what a break up is and what a break up's emotional valence is.

Nagel does not comment on whether extrapolation can be understood as a physical process or not. However, if we are able to understand what some experiences are like just based off their description, and if different experiences have different neurological correlates, then these descriptions might have some neurological basis. Churchland believes our current psychological concepts will be replaced in the face of empirical investigation, but here we may have a way to empirically validate how we currently describe our experiences and ourselves. Let us say that we want to explain how we come to believe there is a relationship between the event of a break up and the experience of heartbreak. Maybe after some break ups you did not feel anything; maybe after others you felt different types of heartbreak. Despite this variation, we make a general claim about break ups by extrapolating from our own case. Then, to discover how extrapolation explains this relationship, we begin to empirically study these two variables, break up events and heartbreak experiences. We can imagine, a la Churchland, an empirical investigation into how we neurologically process extrapolation. It might look something like this: when someone says to you, "I am feeling X," your attention focuses on the memory of a past experience of which X is the name, and if you have both

accessed X, then you have succeeded in extrapolating from your own case. When we consider what we do when we try to empathize with another person, this explanation seems appealing. You would only succeed in recreating another person's experience if you felt what they felt. This is perfectly in line with Nagel's argument, too. If we feel feelings because our brains are in certain states, then if you feel what another person feels there must be some physical similarity between your two brains, which presumably we could empirically determine.

If this is possible, then our conscious awareness of our experiences may have some empirically verifiable content. But does this mean that neurological correlates could corroborate what words we must pick in order to successfully empathize with each other? Could they corroborate which words will allow us to extrapolate so that we can understand an experience we have not undergone personally? Do we want to claim that certain words, phrases, or expressions can have some objective connection to our experiences and those experiences' neurological correlates?

If I ask you whether break ups involve heartbreak, the model suggested above might be appropriate. You might take the question personally, thumb through your experiences, and give an answer. But if I ask you whether sadness involves a negatively valenced emotion, it seems less plausible that your answer would be based off your own experiences. We all know that sadness is a negative emotion, but this is part of sadness' dictionary definition. Would Nagel respond that we could only know what this experience feels like if we underwent it firsthand? Might he even say that the dictionary definition only made sense on the basis of firsthand experience? We can imagine, of course, thumbing through the dictionary and finding a new word that referred to an

emotion you had never heard of before. You would not know what the state that the word referred to felt like. At the same time you would know what the word meant, and you could use it in a sentence. Are our words' meanings totally different from the experiential states they refer to, or is there some intrinsic connection?

Churchland might argue that both our words' meanings, as well as what these experiences feel like, corroborate the same misleading account of how we function. Even if Churchland agreed that there was some necessary connection between what our experiences are like and their neurological correlates, he might not care. He might cite work like Nisbett and Wilson's, and argue that validating the psychological concepts we are currently conscious of is very different from validating that these concepts are the best way to describe ourselves. It does not matter that our current psychological concepts are meaningful or that they have dictionary definitions. These concepts will still be eliminated and replaced because we can do a better job at predicting and explaining how we function. Churchland is not worried about whether his new concepts will be meaningful. We can empirically determine that they will be more useful, so how could it be possible that they were meaningless?

#### Section D: Peter Hacker's Psychological Concepts

Peter Hacker believes that meaning is intrinsically bound up in how we use our language, and that no new language will be able to replicate, much less replace, the way we currently are able to meaningfully communicate with each other. To put this in Churchland's terms, our psychological terms and folk psychological rules do more than predict and explain how we function. No matter how thorough our empirical research, there is something about how we currently describe ourselves that is not replaceable. We

will discuss what this is, and how scientists should investigate our psychological concepts, in the next chapter. His and Bennett's response to Nagel, however, should show us how peculiar their linguistic approach is to philosophizing about empirical investigation.

Imagine that you ask me what it is like for me to undergo sadness at this one particular moment, and I respond, "Well, of course it's a negatively valenced emotion, but I don't really mind; it feels good to feel this way right now." This would be an unusual answer, but it would not be nonsensical. I would have answered your question, even if you would have to interpret how my answer was a description of my sadness. Therefore, Bennett and Hacker redefine Nagel's conception of what an experience is. An experience is "the possible predicates of subjects of attitudinal predicates" (275). In other words, when you ask me what my experience of sadness is, I will not answer that it is a negatively valenced emotion. A negativity is not what I experience. When we describe experiences, we are not describing an entity's characteristics but rather our reactions to some feeling, event, or whatever else. Our reactions are attitudes towards feelings, events, and so forth, reactions that can be pleasant or unpleasant, interesting or boring, wonderful or dreadful.

It seems that Bennett and Hacker have replaced what Nagel thought were experiences themselves with reactions to those experiences. Of course I can have a boring, pleasant or wonderful sadness – you could not exactly stop me from doing so. At first blush, this seems to have completely missed Nagel's point. However I react to my sadness, I am still reacting to one emotion, sadness, and I can still ask what that one emotion is like independently of any reaction. Bennett and Hacker's point, though, is that



you will not be able to answer the question the way Nagel anticipates. Our awareness of what an experience is like will not serve to define or describe the experience in a way that is true for other people or even for one person at separate times (278). If we can have a positively valenced sadness, one that is pleasant, for instance, then the one experience we have in which our sadness is a negative experience does not define sadness. Although we may still use the word sadness to refer to some particular type of experiences, we do not know how the word does this because of having undergone experiences. Rather, Nagel's formulation of an experience as what it is like for an organism to undergo experience X describes how it is we come to empathize with another person's experience (280). Hacker and Bennett believe that when we ask, "What was experience X like for you?" we are asking for explication, not definition. We can only describe experiences by describing our reactions to them.

Nagel would happily agree that it is very hard, perhaps even impossible, to articulate what our experiences are like. But the neurological correlate to unpleasant sadness would still differ from pleasant sadness. We are repeating Nagel's argument – neurological correlates suggest the possibility of an objective, empirical way to describe what seems essentially private and subjective. In contrast, Bennett and Hacker argue that neither reference to the experiences we undergo or those experiences' neurological correlates can define a word. Neither philosophical project can explain why a word is used meaningfully, nor is their reason why complicated. Both are just not how we use language. Maybe we know what red is by virtue of having a sample before our eyes, but we would not know how to describe the brightness of the color, for instance, if that was all we knew about red. And we certainly do not know how to talk about sadness just on

the basis of having experienced sadness ourselves before. According to the way in which we do know how to currently use psychological concepts, we do not know how apply those concepts to anything but humans. Bennett and Hacker's contention is that the rules that govern how we communicate right now also govern what empirical results make sense. If empirical results do not make sense, then Bennett and Hacker call the results nonsense.

I'm trying to describe this as a surprising position to take – how could it be that our language governs our science? – but their position is not quite so grandiose. Think about the difference between a neuroscientist referring to a neurological state as sadness as compared to claiming that the neurological state was one that causes a person to feel sadness (83). Can the first be taken literally, or is it a metaphorical way to discuss the second? Is there no physical thing or physical process that we can point to in the brain and say: that thing is sadness?

But when Bennett and Hacker accuse scientists of speaking nonsensically, they do not mean to say, for instance, that we cannot find a complete mechanism to explain some psychological state or process. Sometimes when we identify a psychological concept, like an emotion, with a neurological state, we may be ignoring other neurological steps required to trigger that emotion. Further empirical research will not resolve the philosophical problem of identifying a psychological concept with physical state. Empirical concerns are a separate matter; all Bennett and Hacker want to do is make sure that we describe how we neurologically function correctly. To do so, and to at all use our psychological concepts in conjunction with neurology, we have to talk about our psychological concepts according to the rules by which we use them now. We know what

it means for humans to be sad, but we do not know what it means for part of a brain to be sadness. This is what Bennett and Hacker call the mereological fallacy: whenever we say that a part of brain is a psychological concept, we have used that psychological concept meaninglessly. When we predicate psychological terms to things and say, “X is sad,” we cannot make anything we want our X. The word “sad” only makes sense when we apply it to people. We would not know what it meant for something other than a human to feel sad. Non-human things can be cold, for instance, and so can parts of humans, but no part of a human can be sad (72).

If Nagel were to say that neurological correlates only cause us to feel experiences, he would avoid the mereological fallacy. Any empirical claim to identity between neurological mechanisms and psychological concepts can likewise remove the philosophical problem of defining psychological concepts from their empirical results. By switching out the word “is” for “cause”, we can postpone having to answer what our psychological concepts are in order to sensibly present our empirical results. If Nagel does believe that there is any sort of a necessary, objective connection between the words we use to describe our experiences and those experiences’ neurological correlates, then he may make a similar mistake. He may believe that our neurological mechanisms do not just cause psychological states like experiences, but define these states.

Over the course of the next two chapters, I will argue that Hacker’s response to Churchland is ultimately parallel to his response to Nagel. There is no organization within the brain or which we can find through empirical research that will be able to displace our folk psychology. We can of course make and test predictions, and we can of course understand the brain as a physical mechanism. This just will not produce a new

language. Our language contains not just a method for communication, but topics about which we are interested in communicating about. For Bennett and Hacker, we cannot have one without the other. The only concepts we can use to communicate meaningfully about humans are the concepts that we already have.

If empirical work explores new concepts and topics, then it will not describe the concepts and topics that apply to us, to humans. Empirical work, they believe, can only proceed by exploring the concepts we have now. Otherwise, we will not be describing our psychology, but some other creature's psychology. Our psychological concepts are the topics which empirical investigation into how we function must be about. Folk psychology, then, contains a conceptual framework that cannot be eliminated. As Bennett and Hacker put it, "This 'conceptual framework' does not merely constitute our 'conception of what a person is'—it also makes us the kind of beings we are" (375).

Suddenly, Bennett and Hacker are engaged in much more than just correcting the way empirical conclusions are phrased. Their position entails two controversial claims, which we will explore in the next two chapters. Hacker and Bennett, especially Hacker, believes that what makes sense is not just a correct use of our language. By selecting the rules that describe the full breadth of how we use language, we determine who and what we are. Though this might seem peculiar, how controversial and peculiar the claim really is depends upon what language "use" is and how this "determination" takes place. However, their second claim may prevent us from interpreting these words in any noncontroversial way. Post-folk psychological concepts and theories, or even folk psychological theories used differently from how we typically use them (like Nisbett and Wilson's argument), irrelevantly describe how humans function. Even if empirically

validated, Hacker believes that it is these types of claims that are useless and misleading, not our psychological topics.

How could this be? Nisbett and Wilson show us so plainly that, if we think we know why we feel a certain way about something, we may be wrong. We may think some article of clothing is very becoming and so purchase it, but Nisbett and Wilson can show us that we would not have thought that that pair of pants so edifying they had not been the pair of pants closest to the right side of the table. It seems very difficult to disagree here. What grounds could Bennett and Hacker have to say that Nisbett and Wilson are wrong? And yet Nisbett and Wilson make no argument analogous to Churchland; they have no interest in whether a new set of non-psychological concepts could possibly replace our psychological ones, whether our language allows us to communicate in a way empirical work cannot replicate. They are just (very important and Nobel prize winning) scientists reporting the observations they make, just trying to pay their bills without engaging in philosophy. Hacker does not care. He believes that he can determine the limits of sense and nonsense in any empirical work. By determining these limits, he believes that he can define what our psychology is—what makes us the kind of beings we are.

## Chapter Two

### Section A: Introduction

Maybe it seems more plausible that we could improve our understanding of perception by studying it neurologically than we could heartbreak. Unlike perception, in order to be heartbroken we need to have conscious opinions and feelings. It seems like who we are as human beings must be involved in understanding heartbreak. Perhaps our neurology could show us that we are misled by feelings of heartbreak in some way, but it seems like we have already got to know something empirically verifiable about heartbreak. Our perception, on the other hand, seems like it must have rules that any curious person ought to be happy to learn. Why do objects look like they bend under water? The answer does not involve neurology, but it does involve empirically explicating the rules that determine how we perceive objects.

For both Hacker and Churchland, there is not a difference between the psychological concepts of perception and heartbreak. In regards to what these concepts are and how we should seek to understand them, the two are engaged in an all or nothing sort of debate. As the mereological fallacy suggests, Hacker believes that the way we use language has already defined what our psychological concepts are. Any empirical studies that disagree use language in a nonsensical way. What heartbreak or perception is cannot be defined by any corresponding neurological state or process.

This might not matter for Churchland's argument. If the mereological fallacy entails that we can never define our folk psychological concepts by pointing at the brain, then so much the better. After all, Churchland does not think that folk psychology is a good theory of how the brain functions. It is sometimes right, but once we start

empirically verifying how we discuss ourselves, we will discover a far more useful way to do so. The mereological fallacy, on the other hand, applies only to the meaning of our folk psychological concepts. Insofar as we can still point to neurological states or processes, identify them and give them non-folk psychological names, we can understand how our brains mechanically function and develop useful predictions based off that. Churchland's argument would be unaffected. The mereological fallacy may be a linguistic point about how our psychological words currently have meaning, but it is then only a linguistic point. Why should how we use our words have any relevance to how our brain functions? Why should the meaning of our words have any relevance to how useful our non-folk psychological predictions could be?

The point of disagreement between Hacker and Churchland is not so much about whether folk psychology can be empirically instantiated and defended. Instead, they disagree about whether empirical work without folk psychology can actually describe how it is we function. For Churchland, concerned with prediction and mechanical explanation, our folk psychology is irrelevant. For Hacker, linguistic conventions are a vital indication of the way we function. Over the course of this chapter, we will work with Hacker's writings from several sources to develop his alternative account of how we function. To do so, we will also have to visit important arguments from Wittgenstein's *Philosophical Investigations*. In two brief examples, Wittgenstein will show us how meaning cannot be defined via a private "pointing" or reference to what a person can introspectively access. Hacker uses these arguments to claim that this is what makes empirical work into our psychology so peculiar. We can point to how our brains mechanically function, but we cannot point to the rules that describe how we

psychologically function. Nevertheless, we can explicate these rules, and in doing so we can recognize criteria that define correct or incorrect descriptions of how we function, without having to eliminate folk psychology. Thus, Hacker will argue that the way we function which he describes cannot be explained by empirical investigation.

Section B: Why does language have any relevance to empirical work?

When we consider the mereological fallacy, we see that Hacker does not believe that whether we can assign psychological attributes to the brain is an empirical question. When we identify sadness with a brain state, if we left something out of the mechanism that causes us to feel sad, then we could criticize this identification on empirical grounds. If a complete mechanism was given, then we could not, on empirical grounds, criticize scientists for saying that neurological states *are* psychological states. This is not Hacker's point. Is he criticizing them on linguistic grounds, accusing them of using language poorly? We briefly discussed this in the first chapter, but we did not discuss how Hacker's argument is more significant than replacing the word "is" with "causes". Hacker does not believe the mereological fallacy is fallacious because he is nit-picky about how he wants scientists to use language. Rather, Hacker believes that the fallacy is a conceptual and philosophical error. He explains that,

One cannot investigate empirically whether brains do or do not think, believe guess, reason, form hypotheses, etc.... until we are clear about the meaning of these phrases, and what, if anything, counts as a brain doing these things and what sort of evidence would support the ascription of such attributes to the brain. ("Philosophical Foundations" 71)

Here, we can see that Hacker has more in mind than just using language correctly. Using language correctly is important because it is the only way in which we can use our psychological language literally. Nagel's argument comes to mind: he believes that psychological, mental states must somehow be physical, neurological states, but that we



currently have no way to understand how this is possible (Nagel 437). If we talk about physical states causing mental states, instead of being mental states, then we get to postpone this philosophical problem in order to clearly present our empirical results. Despite that, does this really mean that the word “cause” more accurately represents what is occurring in the brain, that the word “cause” is a more literal choice?

Some neuroscientists and psychologists have claimed that the ascription of psychological attributes to the brain is meant in a technical sense, that these uses are analogical extensions of our current uses, or that these uses are metaphorical stand-ins for mechanical processes (“Philosophical Foundations” 75). In other words, when scientists say that some suite of neurological mechanisms is perception or heartbreak, they do not mean it literally. Hacker’s response to each objection ultimately boils down to this: if you use psychological attributes in some special sense, then your explanation of that attribute will also suffer from this same sense. If you use the words “perception” or “heartbreak” in some technical sense, then you will not explain how the brain functions until you explain what that technical sense is, and the same is true for analogical extensions or metaphorical stand-ins (77). When Hacker urges scientists to provide a fully mechanical description of the brain instead of using psychological words in any non-literal sense, he is not urging them to more completely describe neural mechanisms. He is not worried about anything physical being left out. He is worried about our psychological concepts being misconstrued, as though some part of their meaning will be left out.

So, finally, we will have to talk about what constitutes a word’s meaning. Unlike Nagel, Hacker believes that we do not extrapolate or infer when to correctly apply psychological descriptions to other people. When someone stubs their toe and begins

hopping on one foot, we do not need to imagine what our own experiences of stubbing a toe is like in order to know what the other person is feeling (although we might if we wanted to empathize with the person and feel, to some extent, what they are feeling). Nor can we only guess, and not know, what a person is feeling because their experiences do not happen to us. Think back again to how a dictionary definition lets us know how to use a word. Ideally, the definition teaches us how and when to use the word. This is what it is to understand a word's meaning, and Hacker believes that other people's behavior teaches us how to use language in the same way. Other peoples' behavior determines what counts as "logically good evidence" for the ascription of psychological predicates to other people (82). This is evidence that is different from empirical evidence. We do not need to empirically verify that smiling indicates happiness and frowning sadness. We could if we wanted to, but that would not explain why these facial expressions communicate. Cursing, groaning and hopping on one foot do not just indicate pain from a stubbed toe. These behaviors constitute the meaning of what it is to have stubbed your toe in the same way that an entry in a dictionary helps to constitute a word's meaning when you do not know the word.

Consequently, there is nothing about having a brain that justifies the application of psychological attributes. This is likely what Wittgenstein had in mind when he wrote, "Only of a human being and what resembles a living human can it be said: it sees, is blind; hears, is deaf, it conscious or unconscious" (104). This is why it is not just bad empirical work to describe brains by using psychological attributes. It is senseless to talk about our brains this way because we have no criteria to judge whether these attributes

are being used correctly. It is not empirically right or wrong right to talk about brains like this. It just does not communicate correctly.

We will have to turn to Wittgenstein to understand why communication requires mutually shared criteria to be successful, and we will do so shortly. Before we move on, we should note several things. Although Hacker and Wittgenstein use the word behavior to emphasize that we can only make judgments about other people's mental states by observing them, behavior has no special role over and above mental phenomena other than letting us assign psychological descriptions to other people. Other peoples' behavior allows us to talk about what we cannot see. Behavior does not redefine what is mental; it allows us to discuss what is mental in other people ("Philosophical Foundations" 82). Moreover, these behaviors often are so closely associated with the mental phenomena they define that we do not think of the two as distinguishable. If someone laughs, then we do not guess that they think something is funny. In contrast, if someone stubs their toe and gives no reaction, then they may be suffering silently and trying to not make a scene, or trying to impress upon others how stoic they are. Here there is more wiggle room for how to interpret behavior. Still, a link exists between behavior (or lack of behavior) and how we will label a person with a psychological description.

By using behavior, however broadly construed, to define psychological concepts' meanings, has Hacker just changed the subject? What is the relevance of the meaning of our psychological concepts to how we understand the way the brain functions? Hacker believes that we cannot separate the two questions. In his view, "Clarification of the psychological concepts that are deployed in psychological investigations is a prerequisite for posing fruitful questions amenable to experimental methods" ("Relevance of

Wittgenstein's Philosophy of Psychology" 2). We can only answer how we psychologically function once we have fully understood what the concepts that we are trying to empirically investigate are. This clarification requires that we understand the criteria that define our psychological concepts and give them meaning.

But this entails that we understand how we function in a way very different from how Churchland imagines we function. Instead of pointing to the neurological states and processes we can observe, how we function is defined by the way criteria establish the meaning of our psychological concepts. These criteria can be concretely stated. As Hacker phrases it, the use we make of our language is a "rule-governed practice" (4). Once we make these rules – these criteria – explicit and understand what behaviors constitute logically good evidence for what psychological concepts, then we have identified the subject matter of empirical work into how we function.

This is the method of the *Philosophical Foundations of Neuroscience*. Hacker and Bennett take a certain field of psychology and neuroscience, such as sensation and perception, volitional movement, or even conscious experience as a whole, and describe the rules around how we discuss that topic in our everyday language. For instance, when we discuss consciousness, we need to distinguish between transitive consciousness, an awareness of something, compared to intransitive consciousness, our consciousness which resumes when we wake up from a dreamless sleep ("Philosophical Foundations" 261). To that end, they argue that consciousness cannot be identified with mental states, but that our concept of what consciousness is is actually broader than our concept of what mental states are (267). This is just one distinction in a long list of linguistic clarifications which Bennett and Hacker present. But these two points alone provide the basis of their

critique of empirical investigations into consciousness that treat consciousness as a single phenomenon. The argument applies equally to philosophical investigations, like Nagel's, that give consciousness a similar treatment.

Later in this chapter we will more fully explore examples of this type of reasoning, both from Bennett and Hacker as well as a pair of authors who have adopted their method. For these thinkers, psychology is a peculiar topic to empirically explore because, unlike some other sciences, its subject matter cannot be given an ostensive definition. Not only can we, obviously, not point with our fingers to what heartbreak, attention or consciousness is. We also do not know what these concepts are by virtue of any private analogue to heartbreak, attention or consciousness. This is in direct contradiction with the central point of Nagel's argument. We do feel anger, we pay attention, and we are conscious. But we do not know the meaning of these phrases by virtue of undergoing these mental states or processes ourselves. Rather, we master the use of these psychological concepts by virtue of understanding the public criteria we share with everyone else for what that emotion, ability, experience, and so forth actually is. Everyone's experience is someone's, but it does not follow that introspection is some form of inner, mental perception just like our normal perception of physical objects (88). Indeed, to describe a psychological term as a name of an experience, a mental state, or even the name of a theory, would not explain anything about the term's meaning. Other people must have grounds for justifying their claims about me, while I can avow claims about myself, but it does not follow that I have direct access to my psychological content which others do not have access to (92). Of course we can visualize red if we want to, and when you do it, you have a private experience. But when you talk about how bright

red is, and when we all agree, that is because of shared conventions around how red is described. When Nagel talks about undergoing an experience, he is describing undergoing distinctive phenomena that have discernable qualities independently of our criteria. His argument centrally rests on there being a “subjective quality of experience”, an aspect to experiences that we are directly aware of and that exists independently of our objective, shared criteria (Nagel 436). Some people might say that they know the difference between emotions because they have felt both at different times in their lives. Hacker might respond that these people have confused knowing the difference between experiences and having felt those experiences in the past. Hacker might respond that while it is true that you have felt both happiness and sadness in the past and that they felt different, you do not know how to describe this difference on the basis of having felt both.

Meaning without criteria is as central to Churchland’s argument as it is to Nagel’s. For Churchland, however, it is not our experiences that we can refer to without criteria, but the way that our brain functions. What prevents us from just using language to express how our neurological states and processes function? The meaning of our words? Even if we do use criteria to understand our folk psychology, why couldn’t we be bilingual, and also use a language that had meaning only by referring to our neurological functions? That would seem to be as much a language based on rules as our current set of folk psychological rules. And what prevents this system of meaning by reference, reference without criteria, from eliminating meaning defined only by our criteria? Why are Hacker’s concepts the only concepts that can produce “fruitful” empirical questions?

### Section C: Wittgenstein's defense of criterial meaning

So far, we have a general sketch of the features of psychological concepts as Hacker understands them. The ways in which we intelligibly use language to describe our mental lives and psychologies is guided not just by a shared eagerness to communicate, but by a shared set of expectations about how to communicate. However, Hacker's stance develops a tone of necessity when we look at it in light of Wittgenstein's private language argument. Hacker's argument about how criteria constitute meaning is adapted from Wittgenstein's writing in the *Philosophical Investigations*. Wittgenstein questions the intuitive idea that we can introspectively and privately determine the meaning of psychological terms. Indeed, we should demand further explanation for why Hacker believes that we cannot know what sadness is without other people to share a set of criteria with. We feel sadness, and we feel a very specific feeling when we do feel sadness. Why is this not enough? Why should we need some set of criteria to know how we feel, and why does this criteria have to be public and communal?

To answer these questions, Wittgenstein has us imagine a diarist trying to keep track of the recurrence of one specific sensation. When that sensation rolls around, the diarist will try to associate the sensation with a sign, S, and write down when she had that sensation. We of course cannot point to our sensations, but we would imagine that we could direct our attention to the sensation such that we are later able to remember what the feeling is like. This is to suggest that the diarist could define the sensation by committing to memory the "connection between the sign and the sensation," as Wittgenstein puts it. But we could only remember correctly if we had some criteria by which to say that the memory-sample we had was correct. You imagine red and you ask

yourself how you know that you are right. You could describe its hue or its brightness, but you could only justify why those were accurate descriptions of red unless you had some criteria that defined what red is. Otherwise, as in the case of the private diarist, “One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can’t talk about ‘correct’” (98).

Without some criteria to define when we have correctly associated a sensation with a sign, we would associate sensations with signs arbitrarily—this is the case with psychological concepts. Yet perhaps a person could privately decide on some set of rules to define a concept that applies to how humans function. More to the point, imagine if you were to tell Hacker, very confidently, that regardless of what he says, you do possess the ability to identify your pains without some criteria of what a pain is. After all, you know your pains by feeling them, not by consciously identifying pains by using criteria, public or private. So this response seems very natural, yet Wittgenstein tersely rephrases the question, “Does (this) mean: ‘If someone else had access to my pains, he would admit that I was using the word correctly?’ To use a word without justification does not mean to use it wrongfully” (105). This is a subtle distinction. Our criteria determine whether we have used words correctly or incorrectly. So when you very confidently claim you do not need criteria to recognize your pains, you are not agreeing or disagreeing with any criteria that define what pain is. You are not, for instance, saying that you recognize your pain because of what sort of sensation it feels like, and then going on to describe how it is pleasant or unpleasant, throbbing or aching, and so forth. In effect, Wittgenstein is saying that if you confidently make this claim, Hacker cannot say that you do so incorrectly. But



you are making this claim without any justification –and this is what criteria, whether private and of your own creation or shared by everyone, provide.

Wittgenstein elaborates on this point in a second analogy. Criteria are not just important for matching invisible, psychological content to words or signs. After all, it seems pretty plausible that we talk about psychological content with less justification than we talk about physical objects. We can disagree about and reinterpret psychological content in a way that we cannot when we are talking about physical objects. Moreover, if I say that I know what the word “pain” means independently of the public criteria for what a pain is, then this would apply to everyone as much as it does to me. Perhaps everyone should tell poor Hacker that they disagree with him. We could all tell him that we either do not need justification to know what pain is or that we use our own, private criteria. To grapple with this suggestion, Wittgenstein asks us to imagine that everyone has a box that only they can look into and no one else. Inside is something, called a “beetle”. We can never look inside anyone else’s box and know what their beetle looks like, but, Wittgenstein asks us, “What if these people’s word ‘beetle’ had a use nevertheless?” (106). Maybe we could use the word “beetle” to pick out anything with six legs, or anything black, or anything that is not bigger than a foot around, or so forth for any attribute of beetles that we could come up with. It is by virtue of these rules that we know what is in the box. Of course, it would be hard to describe what is in the box as having six legs if the box were empty, or if what was in the box was constantly changing. Despite that, Wittgenstein still asks us to imagine what it would be like if the word “beetle” still had a use, regardless of what was in the box. In other words, he is asking us how important the box is to our use of the word “beetle” if we still knew the same criteria

that let us talk about beetles. This whole issue of having a box but not knowing what is in it, or not knowing what is in other people's boxes, would be completely irrelevant if we knew the criteria that define what a beetle is. As in the private diarist's case, the privacy of your boxed beetle, or your mental image of a beetle, is not relevant to determining how we use the word "beetle".

This seems strange because, if there is some connection between the word "beetle" and whatever is in the box, then whatever is in the box better have beetle-like features. But even when what is in the box is not constantly changing, and the word beetle does refer to what is in the box, we do not know what the word "beetle" means because it is the name of a thing. We do not know what the word "beetle" means because we can all point to the same thing when we use the word or because we could all together count how many legs beetles have. Even when we are dealing with a physical thing like a beetle, and not a psychological state like pain, pointing to the beetle is only the start of how we begin to formulate the criteria that let us speak correctly or incorrectly about beetles.

Wittgenstein says, "The thing in the box doesn't belong to the language-game at all; not even as a *Something*" (107). Wittgenstein is thinking of language as a game with specific rules, rules that let us know when our words have successfully meant something. Our criteria are not arbitrary – if all beetles did not have six legs then having six legs would not be part of the word meant. Despite that, we cannot claim that when we talk about beetles, we know what they are *because* we have a mental image of them. We could only claim that our private visualization was a beetle because it satisfied some criteria for what a beetle is.

Criteria are no more or less a part of how we define psychological concepts than how we define physical objects. With physical objects, we can judge whether criteria fit by comparing these criteria to physical objects. If we start to find a bunch of beetles with four legs, then we might want to change our definition of what a beetle is. If we did have a beetle in a box that we could refer to whenever we wanted to talk about beetle, it might not be so interesting to talk about our criteria. When we deal with psychological concepts, on the other hand, recognizing our shared criteria becomes much more important. In the same way that we could not compare the brightness of red to blue without some criteria as to what brightness is and how we should go about comparing colors, we could not understand what sadness is or compare sadness to happiness without criteria to do so. Although it is conceivable that we could privately create criteria to define sadness or the word “beetle”, this would not allow us to communicate intelligibly with each other. We could not understand each other if psychological terms were exclusively defined by means of samples (“Philosophical Foundations” 98). Hacker even argues that there is a disanalogy between pointing at a physical object, as when we check to see how many legs a beetle has, and focusing our attention on our feeling of sadness. Concentrating one’s attention is not a kind of pointing (98). Otherwise, we would suggest that we stand in some one certain relation to every psychological phenomenon we undergo. We think thoughts, feel feelings, and remember memories, and there is no reason to imagine that all of these things present themselves to us in a single manner, through our attention. Moreover, there is no analogue to a sample itself. As we have already discussed, perhaps we define colors by the shades they refer to (Hacker allows

this), but we still require some set of criteria that let us say we have described, remembered, or pointed to the color correctly (99).

Before, we said in response to Churchland that whatever our experience of heartbreak is, it must be something – it is not nothing, certainly. If we call the experience misleading or describe it as an illusion, the term at least still has criteria that define it and give it meaning. Whether it is something or nothing now seems irrelevant to what the word means, what function the word has in our language game. Maybe this is not so peculiar; we can talk about unicorns without ever having experienced one. But we still know what physical parts of other animals we could refer to in order to piece together a unicorn. In contrast, Hacker's argument about our psychological concepts is that we are justified in using them independently of our personal, private experience. Undergoing experiences is not relevant to knowing how psychological concepts operate in our language, as we extended Nagel's argument in the first chapter to suggest. Moreover, the same can be said in response to Churchland's argument. The rules that guide how we use sensibly communicate about our psychological concepts cannot be dictated by how the brain's neurological states and processes operate.

### Section C: Prediction and Usefulness

I do not think Churchland would find his argument particularly troubled by this idea. Even if empirical work only provides suggestions as to how we ought to speak, just suggesting new rules by which we can understand each other, it can still suggest better rules than the rules we currently possess. In other words, even if folk psychology is not a theory about how we function, empirical work can still suggest a replacement for this system of meaning. Churchland is certainly aware that empirical work will not provide

this language itself. Previously, we looked at how Churchland imagined folk psychology's replacement taking root slowly among certain groups of people. This new language might not sweep us off our feet, exactly. How a new, empirically validated way of speaking takes hold, whether it will be well loved and well used, is a separate question from whether this new way of speaking better predicts and explains how we function. Perhaps all Wittgenstein's criteria amount to is that communication requires conventions, and it seems strange to say that empirical work not create new linguistic conventions. Despite that, it is still very conceivable that we could slowly integrate these empirical findings into our mundane conception and vocabulary of how we function.

Churchland believes that empirical findings will eventually describe how we function without any psychological, criterially constituted concepts. However, eliminative materialism additionally requires that he argue that old concepts be replaced by discussion of neurological and physiological states and processes. This is why it is so important that Churchland provide us with assurance that empirical work can create new topics that will better predict how we function than our folk psychological concepts. If we are to move beyond folk psychology, then we cannot just provide new rules for how we describe our perception, for instance. We need to redefine how we see, and we will continue to redefine, Churchland thinks, until we no longer discuss perception anything like we do right now.

Therefore, in order for Churchland to take his eliminativist stance, he cannot only depend upon work like Nisbett and Wilson's. As we discussed in the previous chapter, he will also have to argue that post-folk psychological concepts can make sense in a way that has the potential to replace how we currently communicate. However, if

psychological statements have meaning because of criteria, then those criteria do not apply because they explain or predict how humans function. For Hacker, prediction and mechanical explanation do not enter in to how our language and psychological concepts function. The mind is not some single entity whose functioning can be described through one mechanical theory. It is a diverse and distinctive range of powers of which humans are capable (105). Each psychological word's meaning is constituted by how that word describes our behavior. Churchland believes that folk psychology, as a whole, is a theory ill informed about our neurology. Hacker does not think this is relevant. As we will see, because it is humans we are trying to describe, our empirical research will have to describe the topics humans are interested in, ranging from our personalities, our free will, and last but not least, our heartbreak.

Churchland can certainly encourage us to care more about making accurate predictions. But an argument for eliminativism will have to also argue that post-folk psychological concepts can intelligibly replace our psychological concepts. In the *Philosophical Foundations of Neuroscience*, Hacker seizes on exactly this point, questioning whether empirical studies have described us intelligibly, or whether they describe creatures that, through some misuse of our language, no longer resemble humans. In the first chapter, we looked at several possibilities Churchland imagined for what a post-folk psychological language could look like. It seemed hard, or at least strange, to imagine why and how these suggestions would replace our current set of concepts. Hacker's emphasis is different. He is responding to theorists who already believe their studies describe how we function independently of folk psychology. Although it is in bits and pieces, a post-folk psychological future is already here for

Hacker. Thus, Hacker's strategy is to show how these theories are as nonsensical now as they will ever be. He believes he accomplishes this by showing how these post-folk psychological theories do not describe any creature, let alone us. As he writes,

Not only could students of human nature not abandon these concepts and continue to study psychology, but further, if it could be shown that they had no application to a certain creature, it would thereby be shown that that creature was not a person, nor even a human being. ("Eliminative Materialism" 84)

Delimiting the rules that give our psychological vocabulary its meaning is thus essential to what we do and who we are.

Doing so prior to empirical work allows philosophy to play a vital role in what we ought to empirically study by determining which empirical studies are about us.

Philosophy, unlike empirical work, can describe the rules that determine how we use our psychological vocabulary. Empirical work cannot do this, and any empirical work that does must be examined closely and suspiciously. Hacker believes in a stark divide between the two:

No philosophical question can be answered by scientific inquiry, and no scientific discovery can be made by philosophical investigation. Philosophy can reveal the incoherence, not the falsity, of a scientific claim. (Philosophy: A Contribution, Not to Human Knowledge, But to Human Understanding, 15).

This is why Nisbett and Wilson's argument only supports eliminative materialism in a restricted way. Their work shows that empirical observations will allow us to predict how we behave. Indeed, these predictions are obviously more useful than predictions we could make otherwise—we are often terrible at determining why we have behaved a certain way. Nevertheless, whether we can create totally new, non-folk psychological concepts may not be an empirical question. If how our psychological concepts function is a philosophical, conceptual matter, then we can only begin determining how we function

by exploring the rules, the conceptual framework, underlying our use of our psychological vocabulary. Only topics that satisfy this conception of function, the conceptual framework for making folk psychological statements, are adequate topics of empirical investigation.

So Hacker might leave Nisbett and Wilson alone. The pair may limit certain areas where making introspective claims about cause might not be a smart idea, but it is debatable whether they redefine internal, mental causality. Perhaps they get close when they argue that our folk psychological rules are a result of a priori, cultural causal theories. While this might be a useful explanation as to how the rules we have come about, it still would not explain what the rules themselves are, nor would it dismiss the role these rules play in communication.

However, in the next section we will next turn to two areas of empirical investigation within which Hacker and others do identify conceptual confusions. We will look at responses to the representative theory of vision, which some scientists see as an alternative to the concept of perception as “direct realism”, as well as one theory of how this representation takes place. In these two cases, philosophers John Preston and Severin Schroeder (in the first case) and Hacker (in the second) will question the intelligibility of redefining perception. This is exactly one area in which Churchland, in our first chapter, argued we could improve our folk psychology, and bring the way we discuss perception more in line with how we actually neurologically process perception. Hacker, Preston and Schroeder will all argue that this can be done without replacing our folk-psychological topic of perception.



In a third case, we will look at Hacker's response to Libet's research on voluntary movement. Libet makes the claim that, because the neural areas responsible for hand movement show activation prior to neural activity that corresponds to consciously deciding to perform that action, our awareness of deciding to perform voluntary actions does not indicate when actions have been triggered. Consciously deciding to make voluntary actions does not trigger those actions. Here, Hacker will reject Libet's method, critiquing his conception of what a voluntary action is. However, the conclusions one might make from Libet's experiment are actually well in line with the rules Wittgenstein imagines for how we discuss intentionality. I will argue that in this case Hacker does not just critique the claims scientists make. By enforcing his distinction between empirical and philosophical claims, Hacker, for better or worse, treats science as though it can only make very restricted claims about our psychological concepts. If empirical investigations avoid them, as a purely mechanical explanation of our neurology would, scientists are in the clear. But if scientists want to describe our psychologies, our mental lives, then they can only research and describe psychological concepts as they are defined by our criteria. There is no clear way in which Hacker distinguishes between defining psychological concepts and empirically describing them. Hacker rejects Libet's study because it imports a definition of voluntary movement that we do not use, but I will argue Libet's study can and should have relevance to how we use our psychological concepts. In this third example, then, I hope to complicate Hacker's boundary between empirical and conceptual investigations.

#### Section D: Three examples of Hacker's method in response to empirical studies

Preston and Schroeder respond to empirical work that proceeds exactly as Churchland imagines. Our normal understanding of sight must be understood as a theory about how it is that we are able to see. Whatever this process is, we believe that we are directly presented with what is real and “out there” in the world. This is the theory of direct realism, that our visual field contains the physical object itself, and that the “phenomenal object is identical to the physical object” (qtd. in Preston and Schroeder 2). In contrast, a representative theory of perception entails that what we consciously see, our phenomenal objects, are actually representations of what is “out there”. The brain processes the data that it is presented with, and in turn presents a phenomenal object that may or may be an accurate representation of what actually exists. Therefore, by understanding how we process visual input, we can talk about how we perceive the world more accurately than our folk psychological concept of perception allows.

For example, researchers arguing against direct realism talk about “binocular rivalry”. These researchers divided two images to create two new ones by putting parts of both images together, while putting the inverse parts of the same images together to form a second image. One of these new images was shown to the right eye, the other to the left. Participants thus saw half of both images in one eye and the other half in the other eye. These participants reported intermittently perceiving one whole image and then the other in sequence (Preston and Schroeder 255). This is just one example of how direct realism may be misleading, and there are others, but each piece of evidence that Preston and Schroeder look at comes down to a difference between what we see and what is out there. Even when we do see correctly, the argument goes, these anomalies suggest a

machinery that direct realism and our folk psychological concept of perception cannot explain (258).

However, in Hacker's style, Preston and Schroeder do not believe that direct realism is an empirical theory. If we wanted to define visual perception as the internal construction of sensory data into the end product of our visual experience, then there would be no qualities to what we see beyond what is contained in that end product. Preston and Schroeder respond that that is simply not how we use the word "perception". We do not believe that everything we see is perceived correctly. Don't you believe in optical illusions? Of course you do not believe that every time you see something, you perceive it directly—you can always be wrong about what you saw. Optical illusions, however, provide us with no reason to redefine perception. Binocular rivalry may illustrate some quirk about how our brains process visual input, but by understanding binocular rivalry, we do not understand what perception is. The result of direct realism, Preston and Schroeder write, is "the absurd philosophical theory that we are infallibly aware of everything in front of our eyes" (261). Binocular rivalry can describe how we process optical illusions, but it cannot redefine what perception is. We cannot use binocular rivalry to claim that perception is just a misleading folk psychological concept, with a use that can be better fulfilled. Once we consider optical illusions, scientists arguing for a representative theory of perception cannot claim that our concept of perception does not account for binocular rivalry, nor can they claim that direct realism is a good characterization of perception (264).

It is also not possible to redefine perception as a representative process by describing the role representations play in the neurological mechanism that produces

perception. Describing the neurological mechanism that produces perception is only useful if we are trying to settle an empirical matter, not define a psychological term. David Marr's theory of vision accounts for the entire process of perception, but Hacker argues that this does not allow him to redefine perception as representative instead of direct.

Marr proposes that the brain processes sensory data in stages. The brain first creates a sketch of a "primal image", which describes the light in the visual image. It next creates a 2 1/2-D sketch of the image's surface orientations, and finally, with the help of internally stored 3-D model descriptions, the rest of the image is fleshed out (qtd. in "Philosophical Foundations" 143). Perhaps binocular rivalry even occurs in this last step, in which our expectations for what we should see influence what we do perceive in a top-down manner. Marr's theory also contains a further claim about perception: because a visual scene is processed in stages, as each stage is processed, the data that is processed can be understood as symbolically representing the outside world. The last stage completely represents it, whereas the first stage requires the most interpretation, but we can understand the data being processed in any stage as a symbolic representation of the outside world. It is as though there are three separate points in which we can imagine television screens in our brain, each with an increasingly complete picture of what we will finally perceive.

Like Preston and Schroeder, Hacker distinguishes between our concept of perception and the perception that Marr considers. The conception of seeing that Marr engages is not the concept that we use in our daily lives and conversation. If the brain does contain symbols that represent the outside world, then although each stage contains a progressively more complete representation, each stage is as much a representation of

what is “out there” as our final visual experience. Hacker, meddling, asks how we are able to determine that these stages are actually representations of the same thing and contain the same information (145). Marr probably only believes that these are representations of the same thing because the same stimulus prompts these different stages of neural processing. Nevertheless, there is no way in which we can sensibly talk about our one visual image being composed of three different representations of one object. Maybe Marr imagines that these three different images build on one another by superimposing themselves on the previous stage, developing layer by layer. Yet we do not know what it means to talk about seeing a 2 ½-D image. Moreover, we plainly do not see 2 ½-D images. There are not little television screens in our heads. There is neurological processing, perhaps even in stages. Marr, in contrast, interprets his data to claim that each stage represents a visual scene in the same way our final visual experience represents the outside world. However, Hacker argues that the only stage we can meaningfully describe as perception is the last one, and thus that the previous stages cannot be called representations in the same way the final stage can be (145). At this point, the term “representation” no longer makes sense. Marr may have identified three stages in the process that causes us to be able to perceive, but he has not redefined what perception is.

Much like when we were discussing the mereological fallacy, Marr’s theory cannot predicate any physical process with perception. Marr tries to do this by arguing that each stage, including the last, can equally be described as a representation of the outside world. Hacker believes that when Marr claims that there is a symbolic description of a visual scene within the brain, or a representation, that Marr is not using his words

literally (144). If he is using his words metaphorically, analogically, or in a special technical sense, then he will not really be redefining vision. What he means to do is describe a physical process, what he cannot do is replace or displace our concept of vision.

Nevertheless, even if we say that Marr's new definition of perception is senseless, that does not mean that Marr has not identified some important neural processes. Hacker is not critiquing Marr's observations: "We shall not probe Marr's ingenious analysis of the requirements for deriving an image from a light array, an analysis that may well be apt for the design of machines that can carry out visual tasks" (146). Hacker has no problem with the neural processes Marr has observed and identified, but Marr has described our brains as though they were machines that were not parts of human beings. Should Marr be upset about this? It is weird to say that our visual experience "describes" the outside world, and that each of Marr's neural stages also describes the outside world. Maybe Marr is not using language correctly, but is this a fair price to pay for being able to identify perception as a process within the brain? Are there advantages to doing so that might make us want to use language incorrectly, or even to not describe human beings?

Hacker's argument is thus slightly more ambitious and controversial than Preston and Schroeder's. Whereas they argue that binocular rivalry is not really evidence for the representative theory of vision, but instead identifies a type of optical illusion, Hacker argues that Marr's interpretation of his data is just wrong. Hacker's stance, therefore, is really about limitations on what empirical work can claim. His response to Libet's work on volitional action highlights how controversial this stance is.

Libet, in a series of experiments, showed that the activation of neural areas responsible for movement occurred 500ms before participants' conscious awareness of making the decision to move. Conscious decisions were reported being made 150ms before the action (533). Thus, there is neural activity prior to our decision to move that can explain when we will move. For Libet, this decision or volition is very particularly defined. The sort of willing he is interested in is a feeling that arises without any influence from external sources. Participants were instructed to wait for the feeling of volition to arise, act on that feeling, and then report when they first noticed the feeling. Conscious will, then, does not seem to play an active role in triggering our actions. Although we can willfully inhibit actions in the time between conscious awareness of a volition and executing an action, conscious control is limited to vetoing impulses (538).

By Hacker's account, this should seem like a peculiar version of voluntary actions and volition. You might choose to go to the movies and then choose to walk there, but you do not consciously decide to take each step there, much less experience a feeling of volition prior to each step ("Philosophical Foundations" 227). Feelings of volition are neither necessary nor sufficient for an action to be voluntary. Instead, actions that are voluntary are actions that are done in order to accomplish something. Sometimes, we might imagine, we can have feelings of volition because we want to accomplish something. But our concept of voluntariness does not reduce to Libet's concept of volition (229).

When we believe that we caused an action because we willed it, we do not just mean that a moment of action was temporally near a moment of action. Although this may often be the case, it is too narrow a definition to describe all the phenomena that we

consider purposeful, willed action. “This,” Wittgenstein writes, “contains the germ of the idea that the will is not a *phenomenon*” (78). Wittgenstein is making the same point as Hacker. Our will would not be a phenomenon if we did not experience it every time we took a purposeful, willed step to the movie theater. This could support one reading of Libet’s study: because we can observe neurological activation prior to any awareness of volition, moments of volition cannot be used to define whether actions are voluntary. However, Hacker does not think that we can resuscitate and reinterpret Libet’s findings. Researchers studying perception and binocular rivalry thought their experiment was about one thing (neural, visual representations or sense-data), but it is instead about another (optical illusions). Hacker believes that Libet’s experimental design itself is conceptually confused (229). Because he uses a conceptually confused definition of volition to make his conclusions, his conclusions are just wrong. Marr might have identified the process needed to produce a visual image out of a light array, but Hacker is claiming that no useful conclusions about free will can be drawn from Libet’s studies on volition.

#### Section E: Are Libet’s results impossible to interpret?

It should not be a surprise that Hacker disagrees with Libet’s method. Under his interpretation of Libet’s study, Libet is using empirical methods to determine whether a conceptual definition of free will is accurate. He is trying to conceptually determine what our will is, but without investigating any of the criteria by which we know what the meaning of the word “will” is. On one hand, we can define how we function by claiming that how we function is composed of our current psychological vocabulary and the criteria underlying it. On the other hand, it still seems distinctly possible that how we



function should only be determined by reference to neurological processes. Indeed, why could we not say that the rules that define how Churchland's post-folk psychological language has meaning are rules about how words refer to neural states or processes? Earlier, Wittgenstein told us that to use concepts without criteria is not to use them correctly or incorrectly, but rather to use them without justification. So even if criteria do not underlie "meaning by reference", we may be able to explore this idea of what justification is. If the meaning of our psychological vocabulary is the use we make of it, the distinction between sense and nonsense may not be determined by some rigid set of rules, but by a creative interpretation of how to apply empirical investigations to our lives.

On the face of it, this idea might seem innocuous. Why shouldn't we interpret Libet's work anyway we like, so long as we do not misrepresent empirical work? This seems to be exactly what Libet does, though he still uses language peculiarly. Perhaps because of the controversial philosophical element of his work, Libet is very careful to state that his results only apply on a phenomenological level. He restricts his conclusions to be only about the will as a phenomenon, how we experience willing an action. More importantly, he is not just creating a totally new and foreign concept of what willing a voluntary action is. If I ask if you meant to fidget a moment ago, and if you say that you did not feel any desire to fidget, that would seem to answer my question. We often do talk about willing actions as having a moment of volition. This is a simple, often useful way to talk about whether we meant to perform actions. But just because it is useful in one context does not mean it is useful in all contexts, as when you call walking a willed action despite not constantly experiencing volition. In consideration of other ways we

might use the term “free will”, Libet distinguishes between what he calls “‘philosophically real’ individual responsibility and free will” and the type of will his study investigates (538). Is he really redefining what our will is? When Wittgenstein says that our will may not be a phenomenon, he suggests that it is not enough for our volition to temporally precede an action in order for us to have willed that action. Maybe in order to will an action, we have to endorse our actions, and perhaps this endorsement does not need to come before we make an action. Nevertheless, it might be stranger to come across a person who thinks that you can will an action after it has happened than it might to find a person who did not. We probably all think like the latter type of person sometimes. The former type of person we might only be able to find in a philosophy class.

Libet’s study takes one conceptual definition of what our power to will actions is, and then shows that this conception cannot fully account for the mechanics of the brain. Do we have the freedom to conduct empirical investigations like this? This question sounds funny—is someone going to try and stop us? Yet this sort of empirical work is radically different than just explaining and predicting phenomena. It is radically different from using language to only refer to neural states and processes. If we do have this sort of freedom, then criterial meaning is open to being shaped by empirical investigation. Who and what we are is open to being shaped by empirical investigation.



### Chapter Three

#### Section A: Introduction

On the face of it, empirical questions look different from philosophical ones. Augustine, in his *Confessions*, remarks on how peculiar it is that at any point in the day he knows what time it is, and yet he cannot answer what time itself is (qtd. in Wittgenstein 47). Of course the way anyone might go about answering either question would be very different. To say what time it is, you look at your watch. You can point to the answer, and we could, hypothetically, explain why that observation should be considered good evidence of what time it is. This is an empirical question with an empirical answer. If we ask what time is, we are asking a philosophical question. Whatever the word means in its specific context, we need some set of rules, some criteria, to understand how a word is used and what its meaning is.

It is this sort of philosophical questioning and answering that Hacker deals in. Churchland believes that our grounds for using folk psychology are scant and speculative, and once we refer to empirical evidence in order to put forth a new theory that describes how our brain functions, we can improve upon folk psychology. But of course, for Hacker and Wittgenstein, this does not mean that there are no grounds for using our current set of psychological concepts. Wittgenstein's beetle in a box argument shows us that reference itself requires public criteria, and his private diarist argument shows us that these public criteria are essential for communicating about psychological concepts. Hacker continues Wittgenstein's arguments by saying that we can only use psychological concepts normally when we criterially define them. Terms like "volition", "perception", or "personality" only make sense when we understand the complex,

interrelated rules by which we use these terms. Hacker's contention, throughout the *Philosophical Foundations of Neuroscience*, is that some scientists studying these psychological concepts neglect to address the full breadth of how we use our psychological concepts or create their own, irrelevant uses for psychological terms. When these scientists build theories that seem to explain or explain away our psychological concepts, their results may inadvertently depend upon a redefinition of our psychological concepts. These scientists are then no longer describing our psychology or us – redefinition entails describing how non-human creatures function.

As we have suggested before, that might be exactly what Churchland wants out of an empirical science that explains how we function. Maybe we function less like humans do, and more like a machine might. We can only settle this question by building theories based on evidence that we can observe and point to and seeing which theory is best. However, we already have grounds for claiming that our psychological concepts are at least impoverished, if not totally misguided. If we could develop a completely mechanical theory of how our brains function entirely on the basis of observational evidence, evidence that we can point to and ostensibly refer to, why would this be a worse theory to describe ourselves? And if these theories can better predict how we will behave, why should this theory not be a far better and more useful way to describe ourselves?

The subtlety of Hacker's argument is that he does not reject that we can observe neurological processes in order to describe psychological concepts. We can point to and ostensibly refer to the neurological causes of sadness, for instance. This is how Hacker believes many scientists who violate the mereological fallacy can remove philosophical

problems from their empirical research. Empirical work absolutely can usefully describe how we function, Hacker would say, but it will not usefully describe how we function unless it addresses us and our psychological concepts.

Thus, it is not the possibility but the applicability of a post-folk psychological language that Hacker objects to. Rather, Hacker believes that even if a post-folk psychological language were possible, it would have no valuable application to our lives. Post-folk psychological concepts, once science finally reaches them, are articulable – maybe Marr, with his mechanical brand of perception as a representative theory, is on the cusp of a new concept that will not use our concept of “perception”. As we saw in the last chapter, although this non-psychological theory of perception is possible, it just is not relevant to the concerns we humans have about perception.

I imagine Churchland’s response would be fairly aggravated. If we can make better predictions about what humans will do on the basis of concepts not contained within our folk psychology, why should we insist that these predictions are not relevant to human beings? If we need to consider ourselves more like we currently consider machines than we currently consider other human beings in order to not be misled our self-descriptions, then it’s time to change.

Churchland believes that non-folk psychological concepts will replace our current psychological concepts for two reasons. Firstly, his new concepts will be based upon empirically validated explanations of how the brain works. Correspondingly, the first distinction we will look at between Hacker’s and Churchland’s projects is based in how we can articulate concepts that describe ourselves. Hacker believes that we can only successfully do so on the basis of shared criteria, whereas Churchland believes that we

can only successfully do so on the basis of observational evidence. For Churchland, we do not make observations according to shared conventions, but by ostensibly referring to the data we can point at.

Secondly, Churchland believes that from these explanations we will be able to make better predictions about how humans function. Thus, the second distinction between the two is a result of how both justify their respective tool of choice. For Hacker, the empirical results that are useful are the ones that allow us to communicate about what we want to communicate about. For Churchland, what is useful is what allows us to better predict how we will function. These two different conceptions of use lead the two to different beliefs about what we cannot usefully empirically validate about ourselves. Even if we can empirically validate mechanical concepts of how the brain functions, Hacker believes that because these concepts will not engage how we use language, and thus will not engage our psychological concepts, they will neither replace our concepts or usefully describe us. On the other hand, Churchland believes that even if mechanical concepts of how we function will not engage our psychological topics, they will still be able to replace these concepts. Because we can more accurately, and thus more usefully, form predictions about concepts other than the ones we already have, our current psychological concepts can be replaced.

I will disagree with both positions. The concepts that are useful in describing ourselves do not essentially depend on what we can or cannot criterially define or ostensibly refer to. Churchland is right that prediction must play an important role in how we select what concepts apply to us. Sometimes we must pick concepts that are non-folk psychological and are completely neurological or mechanical. I will disagree with

Hacker, who believes that post-folk psychological concepts, even if they could be created, could not be applied to us because they do not engage the way we currently use language, that is, because they do not engage our criteria. At the same time, I will agree with Hacker that mechanical explanations of how our brain functions will not eliminate our psychological concepts. Ostensively referring to how our brain's function will not interact with our psychological concepts unless we begin to see how this reference engages the criteria that define our current psychological concepts. Once empirical work does this, we will have a third option for how to describe ourselves. We can empirically test which criteria, among the ones we have, are actually useful on the basis of prediction, while that prediction is still intelligible within our language.

It is this position, and only this position, that will allow us to truly change how we conceive of ourselves in light of empirically investigation. Wittgenstein believed that any time we ask for a word's meaning, such as the meaning of sadness, volition, or perception, we could only answer fully by providing the rules that describe how we use the word. Prediction, on the other hand, allows us to select the concepts that usefully describe us, not what meaningfully describes us. Prediction does not tell us how to communicate – only criteria can do that – but it can tell us which way of communicating about our selves is most useful and accurate. Although how we will discuss certain topics of psychological interest, ranging from volition to perception to personality, may change, that we will still be interested in those topics will not change. Eliminative materialism, in other words, does not follow. But neither does Hacker's easy separation of empirical and philosophical concerns. Nobody has exclusive access to the criteria that give our words meaning. If these criteria are legitimately derived from the use we make of words, then,



even if we do not know it, we are constantly engaging with these rules whenever we communicate. Empirical work, especially empirical work about how humans function, is no exception.

#### Section B – Criterial meaning versus ostensive reference?

In the previous chapter, I framed the disagreement between Hacker and Churchland as a debate about the best way to describe how we function. However, both of them intend their arguments to also be about how we should not describe how we function. Hacker's argument is not just about the best way to describe how we communicate and understand each other. It is also about the incoherence of concepts that do not describing us using criteria. Hacker believes that by separating empirical and philosophical concerns, we can peacefully account for the interests of both. We empirically determine facts and philosophically determine the coherence of how we present those facts. Likewise, Churchland's argument is not just about the best way to predict and explain what humans do. Although he spends less time doing so, it is vital to his argument that he identifies folk psychology as a theory, a theory with insufficient evidence to its name. By empirically determining facts, he believes we will determine that psychological concepts, along with their philosophical underpinnings, are not supported by the facts, and so we will replace our old concepts with new, factual ones. This clear-cut divide is only useful for Hacker because of the value he places on criterial meaning, and it is only useful for Churchland because of the value he places on ostensive reference. Once we remove both these valuations, it may be possible that a fully mechanical explanation of how we neurologically function will not interact with our

psychological concepts – or rather, they will not interact until we compare the predictive power and usefulness of both.

Hacker means his divide between philosophical and empirical concerns to ease conceptual difficulties, not cause them. Philosophy determines the coherence or incoherence of claims, while empirical work determines the truth or falsity of claims (“Philosophy: A Contribution to Understanding” 15). Philosophy settles issues of definition such that we can understand the extent of how a concept can be used. As he phrases it, philosophy is a “quest for understanding, not knowledge” (8). Empirical work provides us with facts that require comprehension. When Hacker calls empirical studies nonsense, he is thus not claiming that the empirical study lacks factual accuracy. He is claiming that philosophy must play the role of a “Tribunal of Reason, before which... scientists can be arraigned for their transgressions” (9).

In contrast, when Churchland argues that our folk psychological means of reaching conclusions is a theory, he means to highlight how little evidence it has going for it. Libet’s and Nisbett and Wilson’s studies seem to suggest that our awareness of how we function differs sharply from how we actually function. Folk psychological explanations seem imminently falsifiable. But post-folk psychological explanations will not just represent an improvement in the types of predictions we can make. We will finally have concepts that are derived from ostensibly referring to observable phenomena. By naming and labeling the observations we make with names, we can begin to build completely factual conceptions of how we function. For the first time, humans could build theories out of observations without importing any of the linguistic conventions that we have so far used.

The trouble is whether *we* should value only the self-describing concepts that do not import criteria. Hacker does not think that we can search the brain, find something that we can point at, and then name this thing a psychological concept while using the same meaning that the psychological concept normally has. Factual additions to these meanings are important, but they are not additions to the meanings of our psychological terms, and they are certainly not subtractions.

Hacker never says how factual addition can occur without the meanings of our concepts changing. An uncontroversial, hypothetical example might be that, however we define our concept of focus, there may be an average amount of time we visually focus on stimuli, say when scanning a room. This factual addition to our concept of focusing would still allow us to meaningfully talk about focusing for far shorter or longer periods of time than the average. Our criteria around the word would be unchanged. But say we were to find a unit of neural activity that represented one instance of focusing, and if focusing for an hour involved many of these units. This factual addition might change how we use the word “focusing” when we talk about focusing for hour-long periods. Perhaps focusing is not the correct word to use in that context, much in the same way Libet claims we are incorrect when we say that our feelings of volition trigger voluntary movement. Although we all understand each other when we speak in these empirically incorrect ways, and although criteria can explain why we understand each other in either case, we can do better.

Churchland might not imagine a final point in which we completely understand how our brains function, but he certainly imagines stages. It is hard to imagine any more basic a stage than being able to look inside the brain, point to its processes, and name

them. However, this is of course not how science works. It neither is how mechanical concepts within mechanical theories have meaning, nor is it not how any intelligible communication works. Criteria are essential for all communication, not just communication about psychological concepts. Wittgenstein's private diarist analogy might at first seem to suggest that our use of psychological concepts is special. The private diarist cannot point to her sensation in order to recognize it and know what it, despite experiencing it herself. His beetle in the box analogy makes a similar point by showing that private, ostensive reference still requires public criteria. Even if we had all seen the same beetle, we would have to know what to agree upon if we were to successfully describe it.

While this may seem surprising, it is no different than knowing what red looks like on the basis of having seen it, but only being able to describe it's brightness or hue because of our shared criteria for how to do so. Just because we know what red looks like because we have had some first-hand experience of it, we still need conventions in order to talk about red's brightness. Even if we figure out the rules that red is brighter than maroon and darker than orange on the basis of comparing color samples in our mind, intelligible communication would still require a set of rules. We can also say that red is a color with a wavelength between 620-750 nanometers. When we do, it seems that we could take any color we wanted, determine its wavelength, and define red by ostensively referring to the number of nanometers we get. Yet if someone only told you that any light with a wavelength between 620-750 nanometers was red and expected you to understand, they would obviously not make a good physics teacher. At least by my lights, complicated criteria are required to understand how we physically, empirically define

colors. Post-folk psychological concepts will probably require far more complicated criteria in order to be understood.

But of course this is not a weakness to Churchland's argument. Hacker never says that post-folk psychological concepts will not describe us because they will be unintelligible as scientific theories. A post-folk psychological language will presumably have its own standards for what is intelligible communication. Really, we are misconstruing Churchland's argument when we call a post-folk psychological theory its own language. We do not call the theory of gravity its own language. At the same time, Churchland is at the least imagining a new set of words to describe our selves. It is because a post-folk psychological theory will have its own criteria that it will be able to replace, and possibly eliminate, our current psychological concepts.

Hacker's arguments are not requests for empirical, theoretical explication. The value he places in our current set of psychological concepts is a value in intelligible communication about a certain range of *topics*, the ones humans are interested in. Churchland sees in our psychological concepts a range of possibilities for explaining and predicting behavior. This set of possibilities is not very promising, but there are others. If Hacker values certain topics, than Churchland values others, the topics that we can address by ostensibly referencing evidence and constructing theories. For Hacker's argument, these two values are not in conflict until he claims that we should be interested in only describing the ways that humans function, instead of other creatures. Without this further claim, there seems to be no reason that we could not develop new concepts that describe how we function with our increased ability to ostensibly refer to what is occurring within our brains. It just might not be relevant until we knew how to relate

them to our psychological concepts, and this work might be philosophical. The possibility of describing ourselves in other ways is out there and possible. It just isn't us, so to speak. Likewise, Churchland values post-folk psychological concepts because they are part of a theory created only through ostensibly referencing. But it is only because the predictions these theories could make are better than the ones folk psychology could that we would switch. Scientific theories are criteria built out of ostensibly referencing evidence, and post-folk psychology will be built the same way. But it is because these concepts better apply to us that they will eliminate our current concepts. If folk psychology was not a theory, or if our psychological concepts mattered for other reasons than because they predict and explain how we function, then things might be different.

The crux of Churchland's argument is then that post-folk psychology can better predict the topics within folk psychology. Post-folk psychology will not explain our psychological concepts; it will explain concepts that will replace our psychological topics. In Churchland's argument, our psychological concepts will be eliminated, not explained. In Hacker's argument, these must be explained, and so cannot be eliminated. Hacker believes that our psychological concepts are defined by the behavior that constitutes "logically good evidence" for ascribing them to a person. Empirically investigating these concepts may allow us to apply empirical findings directly to our lives. Yet Hacker will get nothing out of dividing up empirical and philosophical concerns the way he does if he does not also make some claim that our psychological concepts are especially relevant to how we research the way we function. In contrast, we can agree that criteria need to be explicated in order to understand empirical results without agreeing that it is philosophers who need to do this. Philosophy might only be

relevant when we are doing psychology and engaging in our current language – but whether or not we will do so in the face of post-folk psychological concepts is the very question we are trying to settle.

If we ignore Churchland's claim that post-folk psychological topics will predict what we care about, and Hacker's claim that any post-folk psychology will be irrelevant to our lives, then investigating post-folk psychological topics might not lead us to eliminate our psychological concepts. A mechanical theory of the brain will not eliminate the criteria that define our folk psychological concepts just on the basis of including ostensive reference in its repertoire of tools to build concepts. Evidence is important, of course, because it will allow us to discuss what is actually happening in our brains when we discuss our psychological concepts. The right kind of evidence might even make our explanations more or less applicable to our lives. But elimination will not happen just because we use evidence. Unless we begin to talk about usefulness and prediction, a mechanical theory could only contradict our current psychological concepts if it contradicted the criteria that define our current psychological concepts. Empirically validated mechanical theories of the brain will have nothing to do with our psychological concepts if they do not engage our criteria. If post-folk psychological concepts allow us to better predict how we function, as we can when a machine gets an input and gives an output, then there might be some tension between certain psychological concepts and certain post-psychological concepts. We will discuss this in the next section. We can, on the basis of prediction, settle which concepts are useful. But this is a further consideration than how the words of our psychological and post-folk psychological concepts have and will have meaning.

## Section C – Prediction and Usefulness

Does any concept that predicts what humans will do therefore apply to humans? Our psychological concepts do a poor job of predicting how we function, so Churchland believes they will be replaced. This position can seem insipid in the context of Hacker's concerns about topics of interest to humans. Consider our hypothetical empirical result about the average length of focusing. If we reveal the finding midway through the study, does the meaning of the word change from the start to end? It would certainly be right to say that by at the end of the study we at least have one new connotation when talking about focus. Hacker's point still seems to stand despite that. Focus is still the topic of empirical interest, even though how we might talk about focusing has changed. It is not a psychological theory on par with empirical findings and so disprovable by empirical findings. If there are topics of special human interest, then we may have to say that psychology and neurology are special sciences in that we play a role in setting their topics.

Churchland's concern about the predictive power of post-folk psychological concepts is its own way to challenge whether we really should value communicating intelligibly about our current concepts. We can communicate any way we want, so why pick criteria that do a worse job at predicting how we function? This seems tantamount to ignoring post-folk psychological concepts and empirical evidence. Hacker believes that human's behavior determines psychological criteria, but scientific theories should be able to better predict behavior without psychological concepts getting underfoot. Someone stubs their toe, and you have evidence to believe that they are not reacting because they



are trying to seem impressively stoic. Based on the way you have observed this person act in the past, they seem to very interested in maintaining an aura of machismo, and so you believe that it is because of this person's character that they behave in a certain way. However, they may only act this way in certain situations, such as when they feel uncomfortable, and they happen to feel uncomfortable around you. If we had a neurological explanation for this tendency, then neurology might still be interested in the topics of special human interest. But if the neurological process that explained this person's unusual reaction could not be attributed to any psychological state or process in them, then it might be a mistake to analyze this person's behavior through psychological concepts. Will post-folk psychology better predict the same behavior that is fundamental to our psychological concepts, or will it predict totally different behavior?

Churchland sees no problem here. How humans function is no different from determining how any output is produced by some input. If you hold an object up and let it go, it will fall, and in order to explain this phenomenon we talk about the theory of gravity. Humans are only different in that there is some theory, a theory without empirical validation, already in place that explains how we respond to stimuli.

We have seen that Hacker does not agree that we can characterize our psychological concepts as a theory in need of evidence. In the second chapter, we looked at his Hacker's response to the predictive power of post-folk psychological concepts. We cannot group our psychological concepts into a single theory, because as we said earlier, we do not use criteria on the basis of any one goal such as making predictions. Rather, each psychological concept is its own topic of interest, with its own goals for empirical

research. If prediction can apply to how we conceive of both heartbreak and personalities, it will do so in very different ways.

Supposing that a fully mechanical theory of the brain could exist alongside our psychological concepts, so long as it did not engage in the same criteria that define our psychological concepts, how might Hacker construe the predictions it could make? Even if Hacker believes that this sort of a post-folk psychological concept would not describe humans, it would still be important for him to answer the question. Hacker needs to find a way for the improved predictive power of a mechanical theory of the brain to be as irrelevant to human life as post-folk psychological concepts themselves are irrelevant to how we currently use language. Indeed, Hacker's belief that post-folk psychological concepts will not describe humans may ring hollow if no post-folk psychological concepts are yet conceivable.

One way in which Hacker could make this argument would be if he argued that both the predictions and the mechanical concepts that underlie them are equally irrelevant to the lives that we lead. These sorts of predictions would be different from the sort of predictions that Nisbett and Wilson's or Libet's study make. These studies take criteria from our psychological vocabulary and show in what way certain concepts can be misleading. However, it is conceivable that a mechanical theory of the brain would predict totally new and surprising behavior. If post-folk psychological theories do not use our psychological concept's criteria, and if these criteria do depend upon observable behavior, then this may well be possible. Consider Hacker's critique of Libet's study again. He critiques how Libet uses language, but he does not critique how Libet identifies volition with a neurological correlate. If he were to instead question whether anyone

could even begin to make this connection or in what ways this makes sense and in what ways it does not, then his critique would look very different. Instead, he accepts that we can neurologically determine temporal facts about volition, but believes this is irrelevant.

Churchland might not mind either way. If discussing our free will as an inhibitory mechanism more usefully allows us to predict how we make decisions, then there might not be a point in intelligibly applying empirical results to our current concepts.

Nevertheless, I believe that Hacker and Churchland have conflated two very different goals. Post-folk psychological concepts may be useful insofar as they address, via prediction, the topics we are interested in knowing about. Our current psychological concepts may be useful insofar as they are the only concepts that can address the topics we are interested in. We need to examine how post-folk psychology would describe us in order to know if this description really would be irrelevant, or if its predictions could more accurately describe us while still relevantly describing us.

Perhaps Churchland could accuse Hacker of overextending his argument. It is one thing to call redefinitions of psychological concepts senseless if we really are only interested in our psychological concepts' original definitions. Hacker may need a much stronger argument to say that empirical predictions that just ignore our psychological concepts cannot influence how we describe ourselves. Hacker claims that philosophy is needed in to order to understand factual knowledge, in order to understand how the criteria that makes scientific theory intelligible interacts with the criteria that define our psychological concepts. We might not need to do this in order to apply empirical predictions to our lives. This may not lead to an elimination of word's meanings – psychological topics like heartbreak, personal responsibility, or personalities will always

have meaning insofar as we will always be able to communicate about them. Meanwhile, the predictions we might be able to make out of mechanical theory of our how our brains operate might be totally different than the uses we have for discussing our psychological concepts. The two might just ignore each other. We can also imagine, differently than Churchland does, that empirically validated predictions might allow us to adapt our psychological concepts in response to which criteria are most useful to discuss. Whether we explore how psychological concepts will usefully describe us in light of empirical predictions or how post-folk psychological concepts will usefully describe us without psychological concepts, it will be on the basis of usefulness that our psychological vocabulary will either change or not change. I believe that this can constitute a research program significantly different from one that Hacker or Churchland imagines, one that is worth articulating and pursuing.

#### Section D - Psychological concepts, post-folk psychological theory, and a third option

Churchland does not believe that elimination has happened yet – there are no examples that we can point to – but that folk psychology’s doom is evident in the empirical studies that we can currently comprehend. When we were discussing how Churchland imagines post-folk psychological concepts taking root in a population, we described how this might not take place instantly. Although empirically verifiable concepts are inherently better than theories that reference no evidence in their favor, this does not mean that these concepts will instantly become a functioning part of our language. It will take time for post-folk psychological concepts to trickle down to the masses. However, they will be so useful that everybody, not just scientists who value

evidence, may potentially replace their old self-describing concepts with Churchland's radically new ones.

First, however, it will have to become apparent to us that post-folk psychological concepts will replace our old ones because the new ones will accomplish the same purposes our old concepts attempt to satisfy. In section B, I discussed the divide between philosophical and empirical questions that Hacker endorses as well as the different value Hacker and Churchland give to ostensibly referencing evidence when building concepts. I suggested that we could imagine a fully mechanical theory of the brain existing in tandem with our psychological concepts. Post-folk psychological theories will not replace psychological concepts just on account of the former referencing evidence and the latter not doing so.

In section C, I suggested that we will not have much of a reason to value Hacker's or Churchland's project over the other until we can test whether one more usefully describes us than the other. Even if we did value one tool for determining the concepts that describe ourselves, what I think we need to do is imagine the moment in which we decide to either stick with our current psychological vocabulary or discard it. If we are going to endorse or reject eliminativism, then we should imagine how it would be for us to at once have our psychological concepts and a fully mechanical theory of the brain to choose between.

We can create tension between these two types of concepts if we do value Churchland's or Hacker's project more. In contrast, we cannot just choose what types of prediction best apply to us. Three types of prediction have been in play during this chapter. We have been discussing post-folk psychological predictions in terms of the

predictions that a fully mechanical theory of the brain will allow us to make. Hacker does not think that these will be useful. Secondly, although we have not explicitly talked about it yet, empirical work that proceeds as Hacker imagines will contain predictions that are descriptive of and not contradictory to our current set of psychological concepts. Thirdly, we have explicitly discussed predictions that are descriptive of our psychological concepts yet contradictory to them, like Libet's and Nisbett and Wilson's studies.

Both Hacker and Churchland reject this third option. For Hacker, this would deny us the ability to communicate as we currently do and so would describe non-human creatures. This is what motivates his response to Libet's study. Hacker believes that Libet's interpretation of his own results is nonsensical. The version of free will that Libet empirically investigates is not the version of free will that we use, so he is not really describing us. The same goes for a completely mechanical theory of how the brain functions. In neither case would these new concepts fit into our language. It does not matter that we have no post-folk psychological concepts handy to check whether this is true. Ignoring our concepts is as nonsensical as contradicting them. We have no way to philosophically comprehend empirical facts, and so we have no way to integrate these findings into our language.

For Churchland, this third option is still not as empirically sound as a set of completely post-folk psychological concepts. A fully mechanical theory of the brain would import no psychological concepts. Any criteria involved in this theory would be completely separate from the criteria that explain how we use psychological concepts. This third option would introduce ostensibly referencing evidence in order to edit

concepts, and so would produce more accurate predictions than folk psychology, but would still not make the decisive leap of constructing entirely new concepts.

With Hacker's and Churchland's distinct interests in mind, picture yourself sitting here with these three options for how to describe our selves. Which should you pick?

Would you agree with Hacker that we could not integrate empirically adjusted criteria into our current set of psychological concepts? When we compare a fully mechanical theory of how we function to our current psychological concepts, it makes sense to say that there are two distinct languages being used because there are two distinct sets of criteria in play. We could say the same in the third case, but the criteria we would be using are not distinct in the same way. After all, they are still criteria about the same psychological concepts we have used all our lives. Unlike a post-folk psychological vocabulary, the topics of interest are the same. We would still care about free will, perception, and heartbreak. The meanings of these words are not in question – if you state that you feel heartbroken, but I have some reason to doubt that you should take yourself as seriously as you do, I would not hear your statement as garbled nonsense.

Churchland never claimed that we would reject folk psychology because its concepts are not meaningful, but because its concepts are not empirically supported or as useful. The concepts in this third option are both. Empirical results that we can ostensibly refer to will allow us to make predictions that show us in what ways we can adjust our psychological concepts. However, perhaps Churchland's argument is a question of degree. Maybe we should really understand this third option as an intermediate stage before we eliminate our psychological concepts, en route to even more useful concepts. In other words, perhaps concepts that have nothing psychological about

them imported into them, and which are constructed entirely on the basis of evidence, will provide better predictions than either our folk psychology or our empirically edited psychological concepts. Perhaps these better predictions will follow because we will have eliminated any psychological language from our self-describing concepts. Even if this were true, it would still be on the basis of creating the most useful predictions that our psychological concepts would be eliminated.

The accuracy of our predictions is a standard for the quality of our self-knowledge. If one theory can more accurately describe the same topic by more accurately predicting how we function, then that concept is a truer description of us. Psychological concepts as Hacker imagines them are not useful by virtue of their predictions, but by virtue of whether we can communicate about them. They do not enter in to this debate about usefulness in the right way. Insofar as they do not allow us to change our self-descriptions in light of empirical work, we should reject his approach to our psychological concepts. Of course Hacker does allow that empirical facts can affect the connotations associated with our psychological concepts. It is only when we try to define concepts that we cannot rely on empirical facts. However, in our third option, empirical results do not determine the concepts that we are interested. They determine which criteria are usefully relevant, but they do not determine the criteria that give meaning to our concepts. Hacker's concern with how we currently use language is a powerful tool to understand the concepts that meaningfully describe us, but not the concepts that usefully describe us.

Thus, whether or not we endorse eliminativism comes down to whether or not our third option can usefully describe us even when it is possible for us to realize the



usefulness of a fully mechanical theory of the brain. To answer this question, we will look at several empirical studies that in different ways describe how we function.

#### Section E – Applying our psychological topics of interest

Mental rotation is a phenomenon, identified first by Shepard and Metzler, in which people take longer to determine whether a similar pair of three-dimensional shapes are the same shape depending on the degree to which these two shapes are differently angled (Shepard and Metzler 701). In their original study, shapes were either slightly different from one another or the same, and were rotated at 20° increments from one axis. The length of time it took participants to respond could be predicted by the degree to which these shapes were rotated, corroborating participants self-reports that they were consciously rotating a mental image of one shape until it matched the other.

We have discussed how we can only describe what red is by virtue of the criteria that explain its properties. Our mental image of red may enable us to know what red looks like and our recollection of sadness may let us know what sadness feels like, but we can only express these properties through our shared criteria. Shepard and Metzler's study suggests that we can, through private, mental "calculations", determine the similarity or dissimilarity between two objects. However, if we asked how we know these two shapes are the same, either through our shared criteria that allow us to express similarity or dissimilarity or through mental rotation, we would be confused. Our shared criteria allow us to express and articulate why the two shapes are the same or different. This is how we communicate, but it is still possible for us to use the empirically derived concept of mental rotation to usefully discuss and predict how we function. Although predicting how long it takes humans to mentally rotate objects may seem only moderately

useful, it may be just one aspect of our spatial awareness. Perhaps we will come to better understand how spatial awareness works, and so reach even more useful predictions. Perhaps we could predict in what contexts we are better at mental rotation, or what types of people are better at it. None of these claims should be rejected because, in order to express them, we have to use familiar concepts like rotation in strange ways. It is strange to say that the time it takes to perform this rotation reflects how long it might take you to physically perform this rotation, as though you were literally performing it in your mind. It is even stranger to say that athletes may have better spatial awareness because they can perform this task better, or that people can be made to perform better on this task when they are briefly trained in a sport before hand (Moreau 83). Just because these ways of describing ourselves are new, they are not wrong, nonsensical, or irrelevant.

Often, psychological and neurological results will be useful because they describe how we function in ways that are surprising. We do not have much use for empirical affirmations of how we already believe we function, after all. If there is nothing to change, then there might not be anything to do in these sciences. Thus, contradicting or replacing the criteria that define our psychological concepts may be a more important research program. In that spirit, might mental rotation, which in Shepard and Metzler's study based on participant's self-report, be less useful than the predictions we could make if we had a fully mechanical theory of the brain? This seems very plausible. Not only might we expect to get far better predictions on participants' response times. We may also expect to determine what other stimuli are relevant for us to perform mental rotation. We might even rephrase mental rotation so that the concept referred to a mechanical process in the brain – our empirically edited psychological concepts might be just as

susceptible to elimination. Our psychological topic of interest, for all that, would not be changed. Whatever we call it, and however deeply the concept of mental rotation may change, broaden, or narrow, it will not stop describing how humans perform the function of mentally rotating objects. The topic of psychological interest is the same.

It seems fair to say that mental rotation is a function that we will better understand once we understand it as a mechanical process. To a certain extent, then, Churchland might be literally right. Mental rotation might replace discussing participant's reaction times as just slower or faster, and a mechanical theory of the brain might have some concept that will replace mental rotation. Maybe that concept will be something like "spatial awareness", although spatial awareness may still be a folk psychological term. Nevertheless, so long as we have some use for predicting how long it takes people to mental rotate objects, our psychological topic of interest will not be eliminated.

In other cases, post-folk psychological concepts may not be able to encapsulate our empirically edited, psychological topics of interest. For instance, in social psychology, the fundamental attribution error is the tendency we have to attribute other people's behavior to their personality, while we attribute our own behavior to situational factors (Aronson, Akert and Wilson 117). When it is snowing outside, it suddenly seems like everyone who is bad at driving decides to go for a joyride, but when it is your car is skidding down a snowy road, you blame the road conditions. This concept has a very clear explanatory target or topic of interest. We are interested in this difference because we want to know the sorts of attributional errors we make.

In principle, the fundamental attribution error is no more or less neurologically caused than mental rotation. Yet, through understanding mental rotation as physical

process, we might find that mental rotation is actually a less useful term than some mechanical term that explained a broader set of functions. Maybe mental rotation is just an artifact of how we mechanically process spatial awareness (although we are still grasping at straws for a post-folk psychological concept). When we consider the fundamental attribution error and then consider the mechanical process occurring at the same time, any new concepts that could arise from the mechanical process will not allow us to replace our interest in determining how we make attributional errors. Our concern with this error does not derive from any empirical concern with how the brain functions. It comes both from our interest in the way we consistently tend to make certain types of attribution errors and from our concern with not making these errors anymore. Here, it seems fair to say that no mechanical concept of how the brain functions will replace discussion of the fundamental attribution error. We care about the concept because we are social creatures who want to be fair to others, who care about attributing or not attributing actions to people's personalities. It seems doubtful that a mechanical concept could do this more concisely or usefully than our psychological concept of the fundamental attribution error can. Even if one could, here it may be even more obvious that the topics of psychological interest to humans cannot be eliminated the way Churchland imagines.

We have already discussed how Libet's and Nisbett and Wilson's studies contradict our criteria while still engaging our psychological topics of interest. Insofar as the fundamental attribution error contradicts how righteously we usually feel our road rage, it might similarly change how we describe ourselves. But we can also creatively construct useful psychological concepts in order to usefully describe how we function. This is the sort of claim Banaji, Nosek and Greenwald make when they argue that we

have solid empirical ground to discuss prejudice as a subconscious, implicit attitude (280). When we generally speak about prejudice, we speak about a conscious feeling of animus towards some group, and when we generally speak about attitudes, we speak about conscious thoughts that a person has some rational justification for. Psychologists often assess implicit prejudice through implicit association tests, which test the association between positive and negative words and a target group or thing. If participants are quicker to group African American faces with negatively valenced words than they are to group European American faces with positive valenced words, then we may have grounds for calling these participants prejudiced.

If we were more reserved about applying our psychological concept of an attitude or of prejudice, we might not apply these concepts just on the basis of an association. Banaji et al. think this would be an empirical mistake. They cite Eagly and Mladinic's study that shows that people have positive attitudes towards women while having negative stereotype about them (qtd in Banaji, Nosek and Greenwald 283). Likewise, human cognition is not necessarily rational, as "decades of research" have shown. Finally, implicit associations can be very good grounds for ascribing prejudiced attitudes to people. Poelhman et al. have shown that negative association can predict a wide range of behaviors, such as

(U)nfriendliness toward African Americans and gay men, rating a Black author's essay negatively, selecting a Black partner, willingness to cut the budget for Jewish or Asian student organizations, criminal sentence strength for Hispanics, discriminating against female job applicants, and physical proximity to Black partner. (qtd. in Banaji, Nosek and Greenwald 282)

This all suggests that we may have very important reasons for describing our psychological concept of a prejudiced attitude as an implicit association.

Implicit associations might predict discriminatory behavior or stereotypical judgments of other groups, and the notion of prejudice without animus or attitudes without rationale may fit within how psychologists discuss human functioning. Despite that, it might be safe to assume that conscious attitudes and unconscious attitudes are not produced by the same neurological mechanisms. However useful it may be to name implicit associations prejudice, such a creative, even political, application of our psychological concepts may be just plain different than the concepts within a fully mechanical theory of the brain. Can usefulness really justify empirically validating whatever self-descriptions we take an interest in?

#### Section F – Conclusion

Were you wrong, all those years ago in 10<sup>th</sup> grade, when you felt sure you were heartbroken? At this point it should be clear that our concept of heartbreak is the least of our worries. Are you sure that you know what your personality is, or does our concept of a personality not usefully describe how you function? I do not believe the two questions are really opposed to one another. You know what constitutes a personality, and you know what constitutes certain personality traits. But just as you are not charitable or unprejudiced because you want to be charitable or unprejudiced, we do not have personalities because we want to have them. If we can make useful predictions based on our concept of a personality, it will be because this concept usefully applies to our lives.

We should not, without any empirical evidence, believe that we can answer empirical questions. In this way our self-knowledge is limited. But by opening up our concept of personality, for example, to empirical investigation, we gain a powerful way to direct our self-knowledge. It becomes possible for us to learn how we ought to apply

our psychological concepts to ourselves. Post-folk psychology suggests a research program that may not be relevant to any functioning that humans care about, even as it allows us to understand how our brains function. Therefore, elimination will not occur as Churchland imagines it will. While how we discuss our psychological concepts necessarily will change in light of empirical investigation, it is not clear whether post-folk psychological or empirically edited psychological concepts will be more useful and relevant to the concerns humans have.

Moreover, it is not clear whether the concerns humans have, our psychological topics of interest, are susceptible to change. In the first two empirical examples in section E, I suggested that this would not be the case. It is hard to imagine humans whom the concepts of mental rotation and the fundamental attribution error did not apply to, even if we one day stop referring to these concepts by the same name. When we study prejudice as an implicit attitude, on the other hand, we do so in order to address a particular problem. Although prejudice is likely no more or less a universal phenomenon than the fundamental attribution error, describing prejudice as an implicit attitude may be particularly useful for describing how our culture engages prejudice. In contrast to how we neurologically function, this may seem like an unstable, replaceable self-describing concept. Indeed, Churchland might claim that if we really cared about reducing prejudice, then only by referring to its neurological causes could we find the best way to prevent it.

Churchland's argument, then, might still apply, albeit in a changed form. Earlier we questioned whether Churchland could really be an eliminativist if he believed that our concept of perception would only be explained better, not replaced. Here there seemed to be a case in which, even if we were "bilingual" and knew how to describe ourselves

psychologically and post-folk psychologically, the concept of perception would be present in either language, as though it were a cognate. The argument against describing prejudice as an implicit attitude would be, then, that even if prejudice is represented in our post-folk psychological concepts, prejudice as an implicit attitude is not. Not all empirically validated concepts may be created equal. The ones that more closely follow the ark of progress towards post-folk psychology will be the most useful, perhaps even the truest to some kind of neurological human nature.

Churchland deserves an apology for my creative characterization of his argument. Perception as a “cognate” may be a trite way to diminish the potential our psychological concepts have for change. Churchland says that folk psychology is a theory that predicts and explains our functioning, but maybe there are certain functions that folk psychology has identified that will not be eliminated. Churchland could easily think that perception is one such function while free will is not.

Nevertheless, it would be remarkable if, once we have a complete mechanism of how perception occurs, no philosophical problems arose. A complete mechanism of how perception occurs may not require that we redefine perception in the same way it might require we redefine free will. I do not think anyone is too worried about the criteria that define how we use the word “perception” changing. Preston and Schroeder have shown us how at least one attempt to do this only ends up improving how we discuss perception. Only once humans stop caring about where objects are in relation to them will we stop caring about perception. Likewise, only when humans stop making decisions will we stop caring about free will. However, in order to improve how we talk about free will, we will



have to label some of the ways in which we currently talk about free will useless. Still, our psychological topics of interest will not be eliminated.

Our concern when we talk about prejudice as an implicit attitude is not whether our prejudice, as a topic of interest, will one day be replaced. Instead, we are worried about whether the psychological concept of prejudice as an implicit attitude is empirically justifiable. It may be necessarily deficient because it is more creative than mechanical a description of ourselves.

We need to consider two things. One is that a correlation between two variables is not more empirical because it can be explained mechanically rather than with psychological concepts. Correlations tell us whether change in one variable entails change in another. Although we must be able to observe evidence in order to establish a correlation, the strength of their interdependence is irrelevant to whether that correlation supports a mechanical or psychological concept. This means that post-folk psychological concepts' predictions are not inherently more empirical than the predictions we could make with our psychological concepts. Secondly, if Churchland really believes that no post-folk psychological concepts are yet conceivable, although there are plenty of mechanical models that explain how psychological states or processes are caused, we might need some way to tell when our neurology switches from describing psychological concepts to post-folk psychological ones. It would be striking if we did not notice.

At one point in the *Philosophical Investigations*, Wittgenstein compares language to a toolbox, and each different word within language to a specific tool (9). We can also think of our self-describing concepts as tools within a toolbox, and their criteria as possible ways of using those tools. We do not yet know what the best manual will be for

how to use the tools in our toolbox. Some tools may have to be used very differently. Because Hacker does not believe these the criteria that define our current psychological concepts can change, his is not be the most useful research program available to us. That said, he is very right that identifying these criteria is essential to knowing how to express our psychological results. Churchland believes our whole toolbox ought to be thrown away, but that would ignore why our tools were useful in the first place. Predicting how function is a very important way to determine how our tools should be applied, but it does not help us identify what problems our tools should help us resolve. Even if all our tools in the toolbox ought to be thrown out, his argument for eliminativism does help us understand why they should be thrown out. We will have to go tool by tool, problem by problem, to understand why certain tools should stay, why others should be replaced. Perhaps some tools need to be thrown out, others modified, and others may just need to have their rust cleaned off.

However, the debate over eliminativism misses all these considerations. Indeed, the critique over whether we ought to describe prejudice as an implicit attitude also misses this point, at least until we can say in what way calling prejudice an “implicit attitude”, instead of referring to its neurological mechanism, is misleading. How we ought to describe ourselves is a more pressing issue than whether this description will be post-folk psychological or not. If we do not take an interest in how we direct our empirical studies, advertisers may be the only ones who will.



## Works Cited

- Aronson, Elliot, Timothy D. Wilson, and Robin M. Akert. *Social Psychology*. Upper Saddle River, NJ: Prentice Hall, 2005. Print.
- Banaji, Mahzarin R., Brian Nosek, and Anthony Greenwald. "No Place for Nostalgia in Science: A Response to Arkes and Tetlock." *Psychological Inquiry* 15.4 (2004): 279-89. Web.
- Bennett, M. R., and P. M. S. Hacker. *Philosophical Foundations of Neuroscience*. Malden, MA: Blackwell Pub., 2003. Print.
- Churchland, Paul. "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy* 78.2 (1981): 67-90. Print.
- Hacker, Peter. "Eliminative Materialism." *Wittgenstein and the Contemporary Philosophy of Mind*. Ed. Severin Schroeder. London: Palgrave, 2001. 60-84. Print.
- Hacker, Peter. "Philosophy: A Contribution, Not to Human Knowledge, but to Human Understanding." *The Nature of Philosophy, in Royal Institute of Philosophy Lectures*. Ed. Anthony O'Hear. Cambridge: Cambridge UP, 2010. 219-54. Print.
- Hacker, Peter. "The Relevance of Wittgenstein's Philosophy of Psychology to the Psychological Sciences." *P.M.S. Hacker - Papers to Download*. Proceedings of the Leipzig Conference on Wittgenstein and Science, 2007. Web. 28 Apr. 2015.
- Libet, Benjamin. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8 (1985): 529-66. Web.

Moreau, David, Jérôme Clerc, Annie Mansy-Dannay, and Alain Guerrien. "Enhancing Spatial Ability Through Sport Practice." *Journal of Individual Differences* 33.2 (2012): 83-88. Web.

Nisbett, Richard E., and Timothy D. Wilson. "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84.3 (1977): 231-59. Web.

Schroeder, Severin, and John Preston. "The Neuroscientific Case for a Representative Theory of Meaning." *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*. Ed. Timothy P. Racine and Kathleen L. Slaney. London: Palgrave Macmillan, 2013. 253-73. Print.

Shepard, R. N., and J. Metzler. "Mental Rotation of Three-Dimensional Objects." *Science* 171.3972 (1971): 701-03. Web.