

KATRINA L. SIFFERD

IN DEFENSE OF THE USE OF COMMONSENSE
PSYCHOLOGY IN THE CRIMINAL LAW

(Accepted 4 October 2005)

I. INTRODUCTION

The criminal law explicitly depends upon ‘commonsense’ or ‘folk’ psychology, a seemingly innate theory used by all normal human beings as a means to understand and predict other humans’ behavior. This cognitive capacity allows humans to postulate that behavior is causally related to mental states such as beliefs and desires, and to predict or interpret such behavior based on attribution of mental states.

To be found guilty, criminal defendants must be found to possess particular commonsense psychological mental states at the time a crime was committed. Criminal verdicts thus depend upon commonsense attributions of mental states; for example, the mental state ‘intent to kill’, which is comparable to a desire to commit murder, or the mental states ‘knowingly’ or ‘purposely’, which are comparable to a belief that a certain act will have certain results.

However, the past half-century has witnessed several serious attacks upon commonsense psychology (CSP). Certain psychologists and philosophers have argued that CSP is a false theory. On the basis of these arguments some concluded that CSP should be eliminated from the law and replaced with some other, more accurate theory of human behavior.

Below I will discuss two major types of arguments that CSP is not a true theory of human behavior, and thus should be eliminated and replaced. These arguments will be applied to use of CSP in the criminal law. First I will discuss arguments coming from eliminativist behaviorism. Behaviorists worried about the role CSP plays in the criminal law claim that criminal verdicts

should be based upon correlations between ‘objective’ descriptions of behaviors instead of mental states. Next I will explore claims coming from ‘ontologically radical’ eliminativism. Such arguments focus upon the claim that CSP in the criminal law should be replaced with a taxonomy of entities coming from scientific psychology, such as neural patterns of activation.

The success of both sorts of claims depends upon their ability to provide convincing evidence that CSP is a false theory of psychology that thus *needs* a replacement in the criminal law. Only then will arguments regarding what sort of replacement is needed be relevant. Below I will argue that eliminativist projects fail to provide evidence that CSP is a false theory, and thus there is no need to seek a replacement theory of behavior for use in the criminal law.

But first I will begin with a short review of CSP itself.

II. COMMONSENSE PSYCHOLOGY

I see a man glancing over at me repeatedly in a bar. After a half hour he comes strolling toward me with a silly nervous grin on his face. I check the bar stools behind me – nope, no Barbie or cheerleader types there, just a couple of middle-aged men. I groan inwardly. ‘He’s coming over to talk to me,’ I think to myself. ‘He’s going to hit on me.’

My interpretation of the man’s actions may or may not be correct. He may walk right past me to a girl I couldn’t see hovering by the cigarette machine. He may work with one of the men sitting next to me and be coming over to greet them. Or he may want to ask one of *them* out. Or, he may act exactly as I predicted, opening our conversation with the ever popular ‘Don’t I know you from somewhere?’

Despite its less than 100% accuracy, this ability of mine to guess what the man might do next is really quite amazing. I am able to make good predictions of other people’s behavior constantly throughout my day, ranging from the banal (i.e., ‘if I give the guy behind the ticket counter money, he’ll hand me a train ticket’) to the more complex (i.e., ‘the coffee barista forgot about my latte’, and ‘that student is spacing out during my lectures, he’s not going to do well on the exam’). These

interpretations and predictions, which entail attempting to know what others are thinking (or not thinking), allow me to better pick my own behaviors and give me a better chance at having successful interactions with others.

This theory that humans use to predict and understand behavior is termed CSP by many psychologists and philosophers. Wilfred Sellars, the first person to develop the idea that humans use such a theory in their everyday interactions with others, came to the idea by thinking about how humans come to know about their own mental states. Sellars argued that the content of such mental states – such thoughts as ‘I’m hungry’ or ‘I don’t like that guy’ – were not simply given to us via introspection.¹ He argued against this ‘myth of the given’ via a thought experiment involving our human ancestors, imagining that such ancestors were limited to predicting and understanding other people’s behaviors by noting associations between behaviors and the things that preceded them. Thus to predict a pain response, they would have had to remember that getting stung by a bee was usually followed with pain behavior, like grimacing or rubbing the painful area.²

This would have been a particularly laborious way to anticipate human behavior, especially as human behavioral responses appear so varied: one stimulus might produce any number of responses, and the number of things that can produce any particular behavioral response in any human appears to be enormous. In order to be useful as predictors of human behavior, a huge number of associations between antecedent and behavior would have to be stored and available for quick recall.

But, Sellars wondered, what if our ancestors learned instead to posit inner episodes as the causes of overt behavior? This would mean that the facts of some situation could ‘trigger’ a certain concept of a mental state like ‘pain’ or ‘fear’, which would then be linked with stored information about likely

¹ Sellars W., ‘Empiricism and the Philosophy of Mind’, in H. Feigl, M. Scriven, and W. Sellars (eds.), *The Foundations of Science and the Concepts of Psychoanalysis*, *Minnesota Studies in the Philosophy of Science* (Minneapolis, MN: University of Minnesota Press, 1956).

² Ibid.

behavioral responses to that mental state. Behavior in the person one wanted to predict may cause attribution of a mental state to that person, which would then allow for prediction of behavior that tended to be associated with the attributed mental state. This method of behavior prediction would allow for *generalization* across situations and persons predicted via use of a relatively small number of mental state categories (as opposed to a virtually infinite number of actual precursors to behavior one wanted to predict).

Sellars further hypothesized that it may have been possible for persons to learn to apply the theory of CSP to themselves. That is, we may have learned to encounter a certain environmental situation – or to imagine one – and to then apply a mental state response to themselves in order to predict or interpret their own behavior. This might allow one to use their imagination to choose to submit themselves to particular situations that would trigger the mental state associated with the best outcomes.³

Sellars' thought experiment was not meant to be an accurate description of the actual historical origins of a theory of mind in human beings. It was instead intended as an argument against the assumption that humans have privileged access to their own mental states, and that they thus attributed mental states to others by applying to others mental states one had introspected in oneself. However, after decades of research into Sellars' hypothesis, the idea that humans use something like a commonsense psychological theory of mind to attribute mental states to others as a means to predict and understand behavior is now widely accepted.⁴

The theory of CSP places mental states in a privileged role in the explanation of human action, where such states are seen as the source or cause of behavior. CSP is thus thought to work in the following way: my CSP utilizes anything I can learn of your behavior (using my perceptual apparatus and information or

³ Ibid.

⁴ See, for example, Carruthers P. and Smith P., *Theories of Theories of Mind* (Cambridge: Cambridge University Press, 1996); Dennett Daniel, 'Real Patterns', *Journal of Philosophy* LXXXVIII (1991): 27–51; and Fodor Jerry, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: MIT Press, 1987).

assumptions manifest in the relevant mechanisms in my brain) to generate interpretations and predictions of further behavior. These interpretations and predictions postulate two general categories of mental states, desires and beliefs. Desires seem to loosely name goal states, like 'wanting' something, 'hoping to have' something, 'needing' something, or an 'intent' to do something. Beliefs appear to loosely name informational states about a thing desired, like 'knowing' or 'thinking' something is in the fridge, or 'doubting' that the thing is in the fridge. So, in the case of the man at the bar, I attributed to him the desire to talk to me, and the belief that if he walked over to me, I might talk with him.

It is important to note that discussions of CSP in the contemporary literature often refer to mental operations in two very different ways (or at two different levels). First, there is the question of what sort of cognitive capacity CSP is. It is generally accepted that CSP is a 'mental' theory or capacity in the sense that it is instantiated in the brain, but there are many questions about how the capacity is manifested in mechanisms in the human mind.⁵ Controversies surrounding this aspect of CSP have to do with the details of how CSP is developed in the brain, the extent to which CSP can be considered 'innate', and how the capacity of CSP is organized (i.e., is it a discrete or 'modular' capacity, or is it a function of something like 'general intelligence'?).

The second type of discussion of CSP generally revolves around whether CSP accurately describes the mental entities that other people like the man at the bar have inside their heads. That is, if it was the case that I guessed the man had a desire to talk to me, and then he came over to talk to me, am I accurately describing some aspect of the man's mental life? Am I correct in claiming he had a particular thing in his head (accurately described as a 'desire to talk to me') that was in some way linked to his overt behavior?

This is the aspect of CSP that is particularly interesting from the perspective of the criminal law. The reason should be fairly

⁵ This doesn't necessarily mean the theory operates consciously: there are very many cognitive capacities instantiated in mechanisms in the mind that are subconscious, such as some aspects of vision (like construction of 2D and 3D images) and the monitoring of body temperature.

obvious: if CSP is a poor or inaccurate theory of what other people have going on in their heads, the criminal law, which specifically uses a commonsense psychological approach to classify criminal defendants, may consequently be doing a poor job of determining who is guilty of a crime.

Critics of CSP doubt CSP's veracity because it appears to describe entities *in our heads* – entities whose properties cannot be directly perceived via the human senses. How is it possible that our concepts of internal brain states are accurate, they wonder, when it seems clear that such concepts were formed at a time when we at best had indirect access to the properties of brain entities or processes? Can we really expect CSP concepts such as 'desire' and 'belief' to accurately describe real entities inside our heads?

'Realists' about CSP, on the other hand, note that CSP has real predictive power. And even though this predictive power isn't 100% accurate, they argue that the *ceteris paribus* phrases necessary for CSP generalizations are similarly necessary in all of the special sciences.⁶ They also claim that we should take CSP's apparent indispensability to human beings as evidence that CSP is real.⁷

⁶ Fodor Jerry, 'Special Sciences (or the Disunity of Science as a Working Hypothesis)', *Synthese* 28 (1974): 97–115.

⁷ Jerry Fodor, for example, argues that we have no alternatives to the vocabulary of CSP if we want our behaviors and their causes to be subsumed by any counterfactual-supporting generalizations that we know about. Talk of mental states and consequential behavior are of irreducibly psychological categories, and therefore CSP cannot be reduced into another type of explanatory framework in cognitive science. Fodor argues that "...the subsumption of the motions of organisms... by the laws of physics does not guarantee that there are any laws about the motions of organisms qua motions of organisms." Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* 9. According to Fodor, if we want to talk about the behavior of humans, and not the behavior of neurons, we're going to have to use CSP.

As we shall see below, however, it is not required that a realist about CSP believe that human behavior can be explained only in commonsense psychological terms. Realists who believe that commonsense psychological terms can be reduced to another type of explanation claim that human behavior may be explained truly using a commonsense psychological vocabulary *and* some other vocabulary (usually one coming from scientific psychology).

If the realists are correct, and CSP is a ‘true’ theory of psychology, the criminal law is correct to rely upon it to generate criminal verdicts. But, as noted above, if eliminativists are right about CSP the criminal law faces real problems. Eliminativists claim that beliefs and desires are not ‘real’ psychological entities, and thus as a psychological theory CSP should be abandoned for other explanations of human behavior that postulate real psychological states. Such eliminativist positions tend to be motivated by a worry that CSP isn’t ‘scientific’ enough. While the ontology of CSP’s main components, belief and desire, is considered by some to be controversial, the building blocks of other sorts of psychological explanations – neurons, cognitive mechanisms, or outward behaviors, for example – is undisputed (although the details of how they work may be). Plus, eliminativists claim that other theories of psychology postulate testable laws of human psychology, whereas CSP does not. That is, commonsense psychological entities do not seem to be the type of entities which instantiate strict causal laws, whereas other psychological entities, such as neurological ones, are. Thus other psychological theories seem to have explanatory power CSP lacks.

Below I will discuss two different types of eliminativism, specifically with the goal of examining whether they provide convincing arguments that the criminal law should *not* be based upon commonsense psychological concepts. In the end, I will argue that neither gives us reason to eliminate CSP concepts from the criminal law.

III. ELIMINATIVIST BEHAVIORISTS

Behaviorism is the view that human behavior is best explained by external environmental causes, rather than by internal mental causes.⁸ As Gilbert Ryle notes, “Behaviorism was, in the beginning, a theory about the proper methods of scientific psychology.”⁹ Behaviorists argued that psychology’s theories

⁸ Moore Michael S., *Law and Psychiatry: Rethinking the Relationship* (Cambridge: Cambridge University Press, 1984) 36.

⁹ Ryle Gilbert, *The Concept of Mind* (London: Hutchinson’s University Library, 1949) 327.

should be based upon direct and repeatable observations that could be subject to the scientific method. However, "...deliverances of consciousness and introspection are not publicly checkable. Only people's overt behavior can be observed by several witnesses, measured and mechanically recorded."¹⁰ Thus behaviorists argue that true psychological explanations cannot be based upon inner mental states.

Behaviorism's most famous proponent was B.F. Skinner, who argued that a scientific understanding of human behavior should seek to correlate present behavior with past stimuli and behavior.¹¹ Skinner's arguments were notoriously convoluted; however, in his article 'Skinner Skinned', Daniel Dennett expends much energy bolstering Skinner's various claims into a comprehensive argument for behaviorism.¹² Dennett concludes that Skinner's argument boils down to the assertion that talk about mental entities – beliefs and desires – is inimical to science.¹³

To support this assertion, Skinner makes various claims about mental states and our ability to have knowledge of them. Skinner's worries focus upon the 'privacy of the mental': the notion that we cannot – at least, given the current state of science – truly *see* the existence of mental entities such as beliefs and desires. He

¹⁰ Ibid.

¹¹ Noam Chomsky, famous linguist and critic of Skinner, sums up Skinner's project in the following way: "What is so surprising [about Skinner's program] is the particular limitations he has imposed on the way in which the observable of behavior are to be studied..." Chomsky notes that "One would naturally expect that prediction of the behavior of a complex organism (or machine) would require, in addition to information about external stimulation, knowledge of the internal structure of the organism, the ways in which it processes input information and organizes its own behavior." But Skinner attempts to explain verbal behavior without referring to the internal operations of the speaker, on the grounds that "...the contribution of the speaker is quite trivial and elementary, and that precise prediction of verbal behavior involves only specification of the few external factors that he has isolated experimentally with lower organisms." Noam Chomsky, 'A Review of B.F. Skinner's Verbal Behavior', *Language* 35, no. 1 (1959): 27–28.

¹² Dennett Daniel, 'Skinner Skinned', in *Brainstorms* (Bradford Books, 1978).

¹³ Ibid.

thus argues that such mental entities can only be inferred (because they are internal events that cannot be directly observed).¹⁴

What concerns Skinner about the privacy of the mental is the way in which this may allow for a certain sort of explanation of the operations of our inner mental world: explanations that some may mistakenly see as 'scientific'.¹⁵ Specifically, Skinner is worried that explanations will be invented that *presuppose* the rationality or intelligence a theory of psychology is meant to explain. To Skinner, the privacy of the mental often leads those interested in using commonsense psychological concepts in their explanation to attribute a homunculous as a means to explain behavior.¹⁶ Attribution of a homunculous is equivalent to postulating a little rational or intelligent entity inside the head as the explanation for why a person behaves a certain way. To put it another way: to say that the man in the bar walked toward me because he wanted to talk to me presupposes the rationality of the man; it assumes that the man already knows that walking toward me is a means to accomplish the desire to talk to me, and doesn't explain how the man comes to know this. Indeed, as Dennett notes: "Use of all mental terms to some extent presuppose rationality."¹⁷

Dennett concludes that the legitimate assertion lying at the heart of Skinner's complaints about CSP is that "...no satisfactory psychological theory can *rest* on any use of intentional idioms..." for their use presupposes that which psychology is meant to explain.¹⁸ This means that progress in the science of

¹⁴ Dennett claims that Chomsky "...takes this to be Skinner's prime objection against mentalistic psychology..." but Dennett rightly notes that "...Skinner elsewhere is happy to note that 'Science often talks about things it cannot see or measure' so it cannot be that simple." Ibid., 55.

¹⁵ As Dennett puts it: "In several places Skinner hints that what is bothering him is the ease with which mentalistic explanations may be concocted. One *invents* whatever mental events one needs to 'explain' the behavior in question." Ibid., 56.

¹⁶ Dennett quotes Skinner as saying: "Science does not dehumanize man; it de-homunculizes him...Only by dispossessing him can we turn to the true causes of human behavior. Only then can we turn from the inaccessible to the manipulable." Ibid., 57.

¹⁷ Ibid., 60.

¹⁸ Ibid.

psychology cannot ultimately appeal to commonsense psychological terms. Instead, truly scientific explanations of behavior must either (1) translate commonsense psychological explanations into other sorts of non-intentional explanations (or, 'reduce' intentional or mental state concepts into non-intentional concepts), or (2) if necessary, eliminate such commonsense psychological concepts (if it is the case that such commonsense terms turn out to be contradictory to a non-intentional explanation and thus are non-reducible).

Skinner, however, single-mindedly fails to acknowledge option #1 as a solution to his concern about the ultimate appeal to intentional terms in psychology. Instead, he assumes that his observations about the privacy of the mental and CSP's presupposition of rationality require one to resort to option #2 – ceasing use of mental states altogether. But this is throwing out the proverbial baby with the bathwater. As Dennett notes, "There is no reason why intentional terms cannot be used provisionally in the effort to map out the functions of the behavior control system of man and animals, just so long as a way is found eventually to 'cash them out' by designing a mechanism to function as specified."¹⁹

Some behaviorists have been subtler in their leap from methodological concerns to the position of elimination. Michael Moore splits behaviorists into two camps: methodological behaviorists, into which he places Skinner, and philosophical behaviorists, into which he places Gilbert Ryle.²⁰ Moore rightly notes that Skinner doesn't want to reduce commonsense psychological terms to another kind of explanation, and thus refrains from discussing his position as a real threat to CSP.²¹ But it hardly seems that Skinner's project should be reassuring to a realist such as Moore, given that Skinner wants to eliminate commonsense explanations as a 'true' explanation of human

¹⁹ Ibid. Of course, commonsense psychological terms need not be translated into functional explanations, the option Dennett suggests: any sort of non-intentional explanation of behavior may serve to scientifically ground commonsense psychological terms. I will explore some of the options currently on offer in later sections of this paper.

²⁰ Moore, *Law and Psychiatry: Rethinking the Relationship*.

²¹ Ibid., 36.

behavior. On the other hand, Moore sees projects such as that of Gilbert Ryle, who does want to reduce commonsense psychological terms, as more threatening. Possibly this is because Moore correctly views Ryle's project as more complete and/or convincing. In the end, however, Moore claims that Ryle's attempt to reduce mental state terms to 'dispositions to behave in certain ways' is unsuccessful, primarily because of Ryle's false assumption that belief must be reduced because we have 'no verifiable criteria for its application.' Ryle, like other behaviorists, means 'verifiable' in the sense of outwardly verifiable by ordinary human observation. But, Moore rightly asks, why should we require beliefs be verifiable in this manner?²² Beliefs may be internal mental or functional states, not easily observed in the normal human manner. But from this fact one cannot conclude that they don't exist. We can't directly, in the ordinary manner, observe an item's molecular structure. However, we don't conclude from this that statements about that molecular structure aren't verifiable.

One might conclude from the above discussion that behaviorism is dead, and that those of us interested championing use of commonsense psychological terms in the criminal law should not concern ourselves further with behaviorist arguments. However, it turns out that a certain brand of behaviorism is still alive and kicking, and has recently been used as a tool by a prominent figure in legal scholarship to critique the use of commonsense concepts in the criminal law. This project is thus worthy of discussion.

A. Posner's Behaviorism

Judge Richard Posner, active and extremely prolific Federal Judge in the 7th Circuit, has resurrected arguments that mental states are of 'dubious ontology' and thus should not be used to explain behavior in the criminal law. Posner does not adopt the traditional behaviorist position wholesale; instead, he attempts to bolster traditional Skinnerian behaviorism with research in the field of law and economics, which argues that legal practices are best characterized as tools for encouraging economically efficient social relations.

²² *Ibid.*, 38.

In his book *The Economics of Justice*, Posner describes the law as a tool to maximize aggregate social wealth, where wealth is broadly construed as equivalent to something like 'human happiness.'²³ According to this view, the court system is a means for negotiating individual attempts to maximize utility or happiness so that the most amount of people in a society end up with the most amount of happiness possible overall.

To understand Posner's position it is important to examine some of the basic concepts used in models of economic reasoning. The most central assumption in traditional or orthodox economics is that human beings are 'rational maximizers' of their individual desires. This means that given a certain desire, a human being will adopt the most efficient means available to satiate that desire. This in turn implies that human beings respond to incentives: given that I have a desire, it can be encouraged or discouraged by making it easier or more difficult for me to pursue that desire. A rational maximizer of personal satisfaction will adjust his actions according to such incentives to accomplish his ends in the most efficient way possible. That is, given that I have the desire to eat ice cream, I will open my fridge if there is ice cream in it as opposed to walking to the grocery store to buy it, because I believe that the first act would cost me fewer resources for the same gain.

Posner's law and economics approach explicitly embraces a consequentialist view of the criminal law: criminal acts are to be discouraged because crimes lower the total social utility of a society (despite the fact that any particular crime might maximize the utility of the criminal). Put another way, law and economics theory argues that prohibiting and punishing one-sided exchanges like theft or rape confers a net benefit on society by increasing the total amount of human utility (defined in terms of something close to human happiness, broadly construed). Crimes may be deterred via analysis of the performance of criminal acts as cases where the criminal is maximizing his utility and promoting criminal penalties as incentives to refrain from criminal acts. According to this

²³ Posner Richard, *The Economics of Justice* (Cambridge, MA: Harvard University Press, 1981).

theory, the relative weight of a given criminal penalty is best set at the lowest level that will stop the criminal from seeing the crime as an efficient means to accomplish the criminal actor's desired ends: given the penalty (and the chances that the actor will be caught and punished), the crime will no longer maximize the actor's utility. Robbing a convenience store, for example, might maximize one's utility if a large benefit is expected in combination with a low amount of risk. However, if I know the store cash register only contains \$20 and that there is a strong chance – let's say 50% – that I will be put in jail for 10 years as a result of my action, the crime will maximize my utility only if I value my freedom at \$10 or less.²⁴

Like other behaviorists, Posner is motivated to replace commonsense psychological concepts in the law largely due to methodological concerns. Posner worries that we "...cannot peer into people's minds, at least not with the clumsy tools of legal procedure, and if we could we are not at all sure we would find intentions, malice, premeditation, or other entities the mentalist language of the law invites us to expect."²⁵ Because mental entities aren't the kind of things a judge or jury can actually directly observe, Posner concludes that attributing mental states cannot be an accurate process. There doesn't appear to be a 'right' answer when one is attributing mental states: that is, what happens when one juror decides he has 'discovered' a desire to kill, and another fails to attribute such a desire? There doesn't seem to be any criteria by which

²⁴ However, under law and economics models the optimal penalty for any given crime is not determined by its deterrent effect alone. Law and economics advocates explore what degree of penalty will have the optimum deterrent effect by considering the cost of imposing the penalty on society. Although it would certainly be a successful means of deterring speeding, the cost of placing speeders in prison for life would be an enormous burden for those who managed to stay out of jail (resulting in a decrease in the society's total utility or happiness). Hence, due to the constraints of efficiency, the law and economics model is generally used to determine the minimum penalty with the desired deterrent effect.

²⁵ Posner Richard, *The Problems of Jurisprudence* (Boston, MA: Harvard University Press, 1990) 177.

dissentation between two jurors regarding the attribution of a mental state could be resolved.

Posner offers the traditional behaviorist solution to this problem. He argues that mental state terms are just placeholder terms that cover our ignorance of certain causal relationships between certain types of behavior. So instead of looking for the mental states, says Posner, why don't we look for items we can actually see? Using 'objective' descriptions of prior behavior and linking these to subsequent behavior via economic analysis will lead to more consistent and accurate results. Posner argues that if jurors are asked to check 'objective' behaviors off of a list generated by precedent to generate a verdict instead of looking for mental states, such verdicts would depend upon questions for which there is a right or wrong answer, for example: Did the defend buy a weapon or not? Did the defendant gain financially from the crime or not?

Posner thus claims that economic analysis will allow us to determine which prior behaviors are relevant to criminal acts by highlighting the relationship between the pairs of acts. Assuming a criminal actor is maximizing his utility when he commits a crime, acts that are the 'rational' means to the criminal end – the acts which maximize the criminal actor's overall utility given his goal of accomplishing a particular criminal end – are highly correlated with the acts society wants to deter and punish. If a judge or jury can find evidence of these behaviors, according to Posner, then they may safely assume that the defendant is guilty of performing a criminal act.

Let's further explore Posner's proposal via an example. Imagine a man is on trial for the murder of a young boy. The body of the boy was found buried in the defendant's garden, but he claims that he has no idea what happened to the boy. Under the traditional criminal law system, the judge or jury would use behavioral evidence introduced at trial to determine if the man had the desire or intent to kill the boy to determine if he was guilty of murder. Posner argues a court should determine what behaviors might have been performed to achieve the killer's maximum utility, given that he possessed the goal to kill the boy: in other words, behaviors which would express the rational maximizing of this goal. In this case the acts that would

most efficiently lead to the achievement of this goal might have been preying upon the boy over the Internet or arranging to commit the crime ahead of time (possibly by buying a weapon), and attempts to hide evidence of the crime (like disposing of the body and murder weapon).

B. Problems with Posner's Behaviorism

Problems with Posner's project can be placed into two categories: first, problems associated with the replacement he offers for CSP, and second, larger problems that have to do with Posner's aim in offering a substitute for CSP. I'll first review some specific criticisms concerning Posner's particular version of behaviorism. Most of these have to do with Posner's formulation of the behavioral descriptions meant to replace mental state terms in the criminal law. I'll then turn to the larger problems with Posner's version of eliminativism.

1. Many behavioral descriptions are intentionally loaded

Posner envisages replacing 'fuzzy' mental state terms with an 'objective' list of behaviors in order to make the criminal law system more efficient and accurate. But what exactly would such an 'objective' description of a behavior – a description without referring to mental state terms – look like? Behaviors like 'buying a weapon', 'arranging a getaway', and even 'hitting another with one's fist' are intentionally loaded descriptions. As an example, let's look at 'buying a weapon' behavior, a behavior that would seem to be an important correlate to criminal behavior under Posner's proposal. One can imagine that 'buying' behavior might be redefined without reference to a mental state as 'exchange of money for good.' One may buy something without exchanging it for money at the moment of purchase, as in cases where one offers a promise to pay in the future by using a check, credit card, or IOU, but Posner may overcome these difficulties with a more sophisticated definition.

However, it seems impossible to come up with an objective definition of a weapon. What makes something a weapon as opposed to just a regular object? Initially one might think a 'weapon' can be defined as something like 'an object meant to cause harm to another'. But this definition directly refers to a

human mental state: a weapon is an object that a human means to use to harm another. To put it another way: imagine I kill my great aunt by feeding her too many apple seeds in the Waldorf salad I make her for lunch everyday. (Apple seeds contain trace amounts of cyanide.) In this case apple seeds act as a weapon, but what makes this the case? Only that I *intended* them to be used as such. How could Posner ask a judge or jury to look for weapon-buying behavior as a prior behavioral correlate to a criminal act *without* asking them to attribute a mental state to the defendant?

2. The on purpose/on accident distinction

The above problem hints at another related problem. As already discussed, the criminal law currently uses evidence of mental intent (*mens rea*) in addition to evidence of a particular criminal act (*actus reus*) to attribute culpability to a criminal defendant. This is partially because there are reasons to treat the *exact same act* differently for the purposes of punishment. The law is designed to punish the act of taking another's briefcase when the person taking it has the mental state or intent to permanently deprive its owner of his property. However, the law will not punish the same act when the briefcase is taken under the mistaken assumption that one owns it. Thus Posner's claim that mental states aren't required because the "...social concern is with the deed (whether impending or already committed) rather than with the mental state that accompanies it" is only true if his approach can categorize the 'deeds' the criminal law is concerned with in such a way that preserves this necessary distinction between intentional and accidental acts.²⁶

Unfortunately, just as Posner's account provides no way to distinguish between an 'object' and a 'weapon', it also often fails to preserve the distinction between acts we view as criminal and ones we view as accidents. Let's imagine a wife-battering case. A judge's or jury's job is to be able to discriminate between a husband's fist connection with his wife's face when he *meant* it to (i.e., when he intended to hit her), and the husband

²⁶ Posner claims he can do this in the briefcase situation, by asking courts to look to prior behaviors indicating planning to commit the theft or gain from the crime.

swinging around in response to a loud noise and his fist accidentally coming into contact with his wife's face. But what if there are no prior behaviors for the judge or jury to look for? It is possible that in a wife-battering case the only prior act relevant to the hitting behavior was a decision to hit.

3. First-person reports of mental states

It is true that in the wife-battering case a judge or jury is going to have a hard time determining if the defendant intended to hit his wife due to the lack of behavioral evidence of the formation of such intention. That is, this determination may be difficult *unless the defendant confesses to forming such an intention after the fact*. This happens more often than one would think, and the current system is setup in order to take into account such first person reports of mental states. (The entire field of police interrogation is based upon this means of attributing mental states to potential criminals.) Posner's proposal, which focuses upon prior 'objective' behavior, completely ignores this important means by which criminal acts are identified via confessions that one possessed a 'subjective' mental state.

For the same reasons' Posner's proposal also fails to preserve the distinction between non-criminal acts and criminal attempts. In one case walking into a bank with a toy gun in one's pocket is an attempted robbery, and in another it is just visiting a bank carrying your son's toy. In some cases the difference between the two acts is going to purely reside in the actor's intent: there will be no prior behaviors to be identified capable of parsing the two apart.

Thus Posner's 'objective descriptions' of behavior often fail to preserve the distinction between criminal and non-criminal behavior. For this reason, Posner's proposal would seem to return criminal verdicts that are *less* reliable than the verdicts returned under the current system: under Posner's account many criminal acts may go unidentified and unpunished (or, innocent acts may be punished).

4. The indeterminacy of behavior with regard to mental states

Even if Posner could come up with 'objective descriptions' of behaviors that are highly correlated only with the subset of

behaviors that we want to deem criminal, his project still fails. Here's why: in any given case, there is no way to know which subset of such 'objective' descriptions is going to count as reliable evidence that a criminal act was committed. As Sellars noted, any single behavior can appear in a huge number of scenarios (preceded by an enormous number of environmental stimuli or prior behaviors). This means that circumstances or stimuli that vary widely may trigger the same mental state attribution, and very similar circumstances or stimuli may trigger different mental state attributions, even where the attributer and the attributee are held constant. For example, the number of different types of behaviors that would cause me to attribute even a fairly specific mental state to a particular friend (such as 'He wants ice cream') is absolutely enormous. And a very slight alteration in behavior – say, the raising of an eyebrow – could cause a change in the choice of mental state attributed.

Because of this complexity it seems clear that it would be extremely difficult to replace the mental state attributions generally made to find a defendant guilty of the crime of rape, for example, with a specific list of behavioral correlates. Could a truly comprehensive list of behaviors be generated to replace the mental state of 'knowingly having intercourse without consent'? Posner indicates this list might be generated by looking to precedent – the record of cases decided in the past – claiming that he could use precedent to pinpoint the behaviors that will correlate with certain kinds of criminal acts in different situations.²⁷ It seems he imagines a sort of precedent computer program that could take the facts of the current case, match them to the facts of past cases and pull up a list of the behaviors that were taken as evidence of different degrees of culpability. But this is an unlikely possibility, as we will discover further below.

It may be possible that a computer could match the facts of any particular case to the facts of past cases. But even if this computer could come up with a list of behaviors relevant to a particular case, this computer isn't going to be able to provide a list of behaviors that are either necessary or sufficient for a

²⁷ Posner Richard, 'An Economic Theory of Criminal Law', *Columbia Law Review* 85 (1985): 1193–1194.

guilty verdict. Even when the facts of a case match exactly to a thousand previous cases, it is possible that in *this* case those facts aren't indicative of intent, or the type of act the criminal law wants to punish. Imagine I was convicted of murdering my aunt with the apple seeds. The next year, a girl in Omaha is arrested in a case where *her* great aunt died after being fed Waldorf salad everyday. Both of us gained large inheritances as a result of the death of our aunts. But maybe in the second case the aunt loved Waldorf salad and the girl truly didn't know that apple seeds could be poisonous in large quantities. In this case the same behavioral evidence that points toward my conviction for murder should not provide ground for conviction. That is, the exact same behaviors that correlated with my criminal act for murder should not be considered sufficient for a guilty verdict in the second case.

When judges or juries are using mental state terms to attribute culpability, one or two will suffice to establish intent when there is evidence that the defendant was seen running away from the scene of the crime with the murder weapon. But 20 pieces of evidence are required for a guilty verdict in hard cases built on circumstantial evidence. For Posner's theory to work not only would Posner have to figure out a way to create a useful (i.e. not enormous) list of behaviors correlated with the criminal act, but he would also have to come up with criteria for determining which combinations of these behaviors necessitated a guilty verdict. In the example given above, the introduction of a defeasor might be necessary to generate an accurate verdict: a possible statement that the defendant made ('Auntie seems sick, I'm worried') or evidence that the defendant ate the salad everyday herself. In this case Posner's theory fails to generate a correct verdict even where the facts of a case exactly match precedent, because he fails to account for the need of a list of relevant defeasors.

Note that even if Posner did acknowledge a need for a list of defeasors, in any given case this list would be almost endless. If the judge or jury had to look for and then dismiss every possible defeasor in each case the criminal justice system would grind to a halt. Thus Posner's proposal seems an extremely inefficient replacement for the current system.

5. The 'standard' economic model

Finally, a less crucial problem for Posner flows from his dependence upon traditional economic analysis. In addition to the above issues, Posner's account is unlikely to return accurate verdicts because human beings aren't actually utility maximizers. This is an argument that can be raised against most economic analysis of human behavior.²⁸ Human decision-making often deviates from the utility maximizing standard used by most economic theories. Indeed, it has been said that if the standard view of 'rationality' advocated by economic models is

²⁸ There are other, more general, problems with Posner's view of humans as rational maximizers. Although Posner doesn't provide his readers with an explanation of the sort of utility maximizing his theory is based upon, it seems that he uses a standard account of reason; one that would derive norms for human reasoning from formal theories such as logic, probability theory, and decision theory. Generally, these norms are then thought of as 'universal principles' of reasoning, and according to such an account what it is to reason correctly is to reason in accordance with these principles. There is little evidence in support of a standard account of human rationality, and plenty of evidence against it. Problems with the standard account include the following: (1) There seems to be no single way to apply the norms of the standard picture to any particular action (Gigerenzer G., Todd P., and ABC Research Group, *Simple Heuristics That Make Us Smart* (New York: Oxford University Press, 1999).), (2) The principles of the standard picture are subject to different interpretations (Gigerenzer, Todd, and Group, *Simple Heuristics That Make Us Smart*), (3) Different formal theories lead to incompatible claims about human reason, and it isn't clear which ones should be used to derive the principles of reason (Richard Samuels, Stephen Stich, and Patrice Tremoulet, 'Rethinking Rationality: From Bleak Implications to Darwinian Modules', in E. LePore and Z. Pylyshyn (eds.), *What is Cognitive Science?* (London: Blackwell, 1999), and (4) It isn't clear that a derivation of these formal theories leads to actual norms of reason at all. For this one would need a 'principled account of the correct conversion schema for rewriting formal rules as normative principles, and there isn't one. (Samuels, Stich, and Tremoulet, 'Rethinking Rationality: From Bleak Implications to Darwinian Modules.')

It is thus a problem for Posner that he isn't clear on which 'standard accounts' of reason he is using. Determining the norms of reason shouldn't be left up to judges or juries – as a result defendants would be subject to different standards of culpability in different courtrooms. But even if Posner did advocate a single account of reason, it won't necessarily translate into specific norms of reason that could be applied to determine guilt.

correct, then the traditional picture of humans as the 'rational animal' may be false. According to Tversky and Kahneman's Nobel Prize-winning research, if held up to the traditional standard humans exhibit a host of systematic errors when reasoning.²⁹ It seems, contra Posner and others of the law and economics movement, that humans often don't maximize their utility by performing the least costly or most rational actions in order to achieve the end with the most net benefit: instead, they often choose actions based purely upon fast and frugal heuristics or rules of thumb.³⁰

It might seem that Posner could defend himself against this argument by claiming that while certainly in some instances human rational fails, *overall* human beings tend to follow the rules of utility (broadly construed) maximization. But as Tversky and Kahneman note, the use of heuristics to make decisions results in actual 'irrational' behavior when judged by the standard model of reason: the cognitive biases they document are not attributable to the mistakes or malfunctioning of a reasoning system, or in the distortion of judgments of payoffs and penalties. That is, the 'mistakes' made in human reasoning appear to be systematic. They aren't just *performance* errors, where the human cognitive system is randomly going awry. They represent a regular failure of human beings to think in a truly rational or efficient manner under certain conditions.

Given that human beings are often 'irrational', it seems that Posner's economic analysis of criminal law could lead courts to base criminal verdicts on evidence of 'rational' behavior humans aren't likely to perform (or behavior humans aren't even capable of). Imagine a defendant plots for weeks to kill his wife by stabbing her repeatedly with a very small paring knife. Let's say there was a loaded gun in the closet that would have killed his wife instantly and created less of a mess, and thus much less evidence for the police to discover later. But the man doesn't use

²⁹ Kahneman D., Slovic P., and Tversky A. (eds.), *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982), Tversky A. and Kahneman D., 'Judgment under Uncertainty: Heuristics and Biases', *Science* 185 (1974): 1124-1231.

³⁰ Gigerenzer, Todd, and Group, *Simple Heuristics That Make Us Smart*.

the gun: he uses a paring knife. In this case did our defendant maximize his utility given his desire to kill his wife? It certainly seems that this defendant committed a less than optimal murder. Stabbing his wife repeatedly with a small knife was inefficient because it took longer and used up more energy than shooting her with a gun would have. Plus, using a small knife made it more likely that the defendant would be caught and punished because the method actually created evidence against him (i.e., blood splattered on the defendant's clothes and shoes, and under his nails). But in this case the less than optimal nature of this crime shouldn't be considered relevant to the question of the defendant's intent to kill. Indeed, while this man might be seen as less culpable under Posner's account (or not culpable at all), he did indeed premeditate this murder, and thus most would argue he is deserving of the most severe punishment for his act.

6. A failure to eliminate, a failure to reduce

Finally, let's back up a step and consider Posner's larger aim of throwing mental state terms out of the courtroom. If Posner sees himself as providing reasons for eliminating CSP from the criminal law, his project fails for the same reason that Skinner's attempt to eliminate fails. Many behaviorist arguments seem rooted in an outdated view of mental entities as a 'ghost in the machine'. Dennett notes that Skinner seems to see the mental as 'immaterial'.³¹ Posner similarly describes the mind as "...not only unobservable but immaterial, yet despite its immateriality it seems to be in control of a material object, the body."³² And Posner, like Skinner, thinks the problem of how we can know about the mental leads to attribution of a humonculous as an explanation for behavior: Posner claims the mind is seen as "...the invisible puppeteer, the inner man and woman."³³ It is this idea of the mind as an immaterial center of decision-making, Posner argues, that should be discarded, "...despite law's emphatic (but I shall argue, shallow) commitment to it."³⁴

³¹ Dennett, 'Skinner Skinned', 55.

³² Posner, *The Problems of Jurisprudence*, 165.

³³ Ibid.

³⁴ Ibid.

One feeling charitable might assume that Posner's concern about attribution of a humonculous is due to the legitimate behaviorist worry discussed above regarding explanations of behavior ultimately resting upon intentional terms. As discussed, this sort of explanation presumes what is meant to be explained by a psychological theory of behavior: human intelligence. However, as already noted, although a psychological theory may not rest solely on intentional concepts, it need not throw them out altogether. Posner is wrong to assume that mental entities *by definition* fail to have a physical description. Philosophers have spent the past 50 years moving the debate about mental entities beyond the mind/body problem of Descartes and into the realm of physicalism and realism. It is now generally agreed that the mind and the body are not separate entities, and thus, in so far as they exist, mental states are real things in the world operating under a physical description as well as a mentalistic one.³⁵

In short, Posner, like Skinner, fails to acknowledge the possibility that commonsense psychological explanations may translate into other sorts of non-intentional explanations. In order to rule out this option, Posner must argue that commonsense psychological concepts fundamentally conflict with scientific psychological explanations of the causes of behavior. And again, Posner, like Skinner, fails to make such an argument. Indeed, by offering re-definitions of mental state terms Posner actually relies upon such terms to act as useful categories for predicting and understanding behavior: he actually attempts to *reduce* them to another vocabulary.

As already noted, reduction does not entail elimination of CSP concepts from the law, because they can be redefined in terms of the concepts of a replacement theory such as Posner's theory of law and economics. Posner's project entails starting with CSP concepts such as 'intent to kill', and then attempting to redefine these concepts in an 'objective' or behaviorist way. Thus it seems that Posner believes that CSP concepts refer to

³⁵ Smart JJC, 'Sensations and Brain Processes,' *Philosophical Review* 66 (1959): 141–156.

something real, but this ‘something’ is better understood using a behaviorist/law and economics lexicon.

Unfortunately, Posner fails to offer us a legitimate reduction of CSP terms used in law, for all the reasons cited above. Posner tries to redefine the mental state terms normally attributed to a criminal defendant as a disjunction of prior behaviors [b1, b2, b3, etc], and then argues that these redefinitions can be used to replace the mental state terms to the criminal law’s advantage. For his project to be successful, however, Posner would need to provide a comprehensive list of behaviors that generally ‘trigger’ attribution of a mental state in a particular criminal scenario. This turns out to be an impossible task, for the reasons stated above. Thus replacement of mental state terms with such ‘objective’ descriptions is similarly impossible, and Posner’s project fails.

IV. ‘ONTOLOGICALLY RADICAL’ ELIMINATIVISM

We can now turn to a second sort of eliminativism about CSP: ‘ontologically radical’ eliminativism.³⁶ Paul Churchland, probably the most prolific and well-known of the eliminativists, has made it clear in his writing that his arguments are offered in support of a project of radical eliminativism. Such projects usually entail an attempt to eliminate and replace CSP with some scientific theory of psychology. Churchland, for example, champions neuroscience as the theory of scientific psychology with which CSP should be replaced. Assuming the truth of neuroscience, Churchland argues that “... [w]e cannot expect a truly adequate neuroscientific account of our lives to provide theoretical categories that match up nicely with the categories of our commonsense framework. Accordingly, we must expect

³⁶ ‘Ontologically radical’ eliminativism has been distinguished from ‘ontologically conservative’ eliminativism in the philosophy of mind literature precisely to distinguish proposals that do not allow for reduction from those that do. See, for example, Savitt S., ‘Rorty’s Disappearance Theory’, *Philosophical Studies* 28 (1974): 433–436. and Stephen Stich, *Deconstructing the Mind* (New York: Oxford University Press, 1996) 94.

that the older framework will simply be eliminated, rather than reduced, by a matured neuroscience.”³⁷

Ontologically radical eliminativism is an extreme doctrine that advocates wholesale rejection of CSP as a theory because it is radically false (its concepts fail to refer). Because CSP is a false theory, radical eliminativists do not feel that CSP concepts will ‘latch onto’ or be translatable into a ‘true’ theory of psychology.³⁸ That is, they don’t feel that CSP concepts refer to anything real, and thus describe no real phenomena that could be also described in another vocabulary. This means that ontologically radical eliminativism, like eliminativist behaviorism, rejects the possibility of reduction of CSP to a scientific theory of psychology.

Let’s try to put such radical eliminativist projects into perspective. Commonsense psychological concepts – those concepts we use in order to understand other people’s behavior – are a sub-set of the commonsense concepts we use to understand the world. For example, commonsense physics applies concepts such as ‘solid’, ‘liquid’, ‘heat’, and ‘motion’ to the world as a way for us to better understand how everyday inanimate objects tend to act in certain conditions. In the course of human history, our ideas about what these concepts refer to have changed considerably. For example, we used to think of solidity as being equivalent to something like ‘no spaces within’. But our beliefs about the objects the concept ‘solidity’ refers to has changed due to our discovery and increased understanding of the atom.

When we discover that the way in which our commonsense concepts understand phenomena is incorrect, we are faced with two options. First, we may determine that the commonsense concept just doesn’t refer truly to a kind or type of object in the

³⁷ Paul Churchland, ‘Eliminativist Materialism and the Propositional Attitudes’, *Journal of Philosophy* 78 (1981): 67.

³⁸ As Brian Loar notes: “...If it were to turn out that the physical mechanisms that completely explain human behavior at no level exhibited the structure of beliefs and desires, then something we had all along believed, viz. that beliefs and desires were among the causes of behavior, would turn out to be false.” Brian Loar, *Mind and Meaning* (Cambridge: Cambridge University Press, 1981) 14.

world. In this case, it would make sense to completely abandon the commonsense concept because we have discovered the referent of the commonsense concept doesn't actually exist. This, in a sense, is what happened with the commonsense psychological concept 'witch'. People used to believe that this commonsense concept truly referred to some object in the world – a woman with magic 'powers' – and thought when they used it they were referring to this kind of object. However, an essential component of the previous concept proved wrong: persons with magic powers don't exist. In modern western societies, the concept now tends only to be used to refer to 'that thing we once thought existed' (a woman with magic powers). The concept is thus no longer used as a way of understanding the world or explaining worldly phenomena (such as why farmer Ted's crops died).

However, in many cases a commonsense concept is still used to refer to something in the world even after our beliefs about the thing (or things) that the concept refers to change. If the thing the concept refers to exists, then the concept can still be used to refer to that thing even if some of our knowledge or beliefs about it turn out to be false. The concept of solidity, above, is one example of a concept where our beliefs about its referent has changed. Our beliefs about the way in which an object is solid has changed, but the class of entities 'solid objects' can still be said to exist, and we still apply the concept of solidity to those objects. That is, our beliefs about solidity have changed, but the concept still refers. And our improved understanding only helps us better use the concept as a means to understand and explain real world phenomena.³⁹

This is why, although our scientific understanding of the world seems to have changed rapidly over the past few hundred years, it is extremely rare that a commonsense concept is abandoned completely. Similarly, it is rare for a single user of

³⁹ This reinterpretation of a referent of a commonsense concept can occur not just due to new scientific knowledge, but due to any new personal knowledge. Embarrassingly, because I watched too many 'Love Boat' episodes at too young an age, I thought the term 'night cap' referred to an adult sleepover until I was 22, when someone finally set me straight.

the commonsense lexicon to stop using a commonsense concept based on new knowledge. Children learn a commonsense vocabulary when acquiring a language and tend to continue to apply these concepts to the world despite advancing knowledge of science and alteration of their beliefs about the things that these concepts refer to. For example, although most people learn that water is H₂O, we continue to refer to it by its commonsense label, water.

It may be that one of the reasons why we tend to hang onto commonsense concepts even though our understanding of the things they refer to change considerably is because such concepts are a particularly entrenched and vital tool used by humans to understand and interact with the world around them. The attribution of mental states appears to be a universal human trait: all normal human beings understand and predict behavior by attributing mental states. Children begin to understand agency, attribute goal states, and attribute belief states via eye direction detection as early as 9 months of age. This capacity is progressively generalized to a capacity to attribute the full range of mental states (including false beliefs) by 4 years of age.⁴⁰ Persons who fail to develop this ability – it is thought that autistics lack precisely this capacity – suffer greatly in the areas of social interaction and cooperation, and are often unable to maintain relationships, hold jobs and accomplish tasks as simple as taking public transportation or buying groceries.

Thus the application of commonsense psychological concepts is an ability that is vital to our functioning in the world from a very early age, and it is a theory that, as one becomes older, continues to be the most important tool we have to understand and predict others' behavior.

It is precisely because application of commonsense psychological concepts is such a fundamental human capacity that a project that aims at throwing out these concepts – ontologically radical eliminativism – should be considered so radical. It seems

⁴⁰ Baron-Cohen S., *Mindblindness: An Essay on Autism and Theory of Mind (Learning, Development and Conceptual Change)* (Cambridge, MA: MIT Press, 1995) 127.

that only if it is clear that commonsense concepts don't truly refer to anything real in the world, like the concept 'witch', is throwing them out be justified on the grounds that we ought not to refer to non-existent objects to explain worldly phenomena.

We have even more reason to be cautious about radical eliminativism when we consider the role that CSP plays in the criminal law. CSP is used by the criminal law to categorize offenders (as guilty, not guilty, insane, etc) and to determine the appropriateness of penalties, some of them, very severe. Certainly, if CSP were an inaccurate theory we wouldn't want to continue to use it, because we would expect the theory to categorize offenders inaccurately.

However, it is difficult to imagine what our criminal justice system would look like if we were forced to eliminate CSP from the criminal law. Even our most advanced scientific theories of behavior, such as neuroscience, are not yet developed enough to offer an alternative taxonomy of entities that could be used to generate verdicts. And even if such an alternative taxonomy were developed, it is still likely that the criminal justice system as currently conceived could use the theory to generate verdicts. In the US, current criminal statutes refer explicitly to defendant's mental states as the primary means for determining guilt. All such statutes would have to be re-written naming alternative psychological states or entities necessary for criminal guilt. Plus, as discussed above, it comes naturally to judges and juries to use CSP to attribute mental state in their attempt to understand a defendant's behavior. It is unclear that judges and juries could (1) 'turn off' this innate capacity to attribute mental states, or (2) become competent enough in a scientific theory of psychology, such as neuroscience, to generate verdicts based upon application of concepts such as attribution of certain neural or chemical patterns.

Obviously, any attempt to eliminate CSP for a scientific theory of psychology would be useless to the current criminal justice system if judges and juries proved unable to understand or apply their theory to return verdicts. It is likely that if forced to use another theory of psychology, the current criminal justice system would have to be dismantled, and a completely new

system of generating verdicts – possibly one not utilizing juries and judges – would have to be developed.

On the other hand, reduction of CSP concepts to a scientific psychology poses a much lesser threat to the current criminal justice system. We may, for obvious reasons, decide not to employ the scientific terminology in every instance to generate verdicts, and we could do so knowing that the CSP concepts we currently use refer to real psychological phenomena. Further, if CSP could be reduced to scientific psychology, new evidence coming from scientific psychology may be seen not as a threat to but as a means for improving upon the current system. Evidence coming from scientific psychology, would provide the possibility of improving our understanding of the referents of the CSP concepts used in the criminal law. As a result, we could begin to apply these concepts in a more accurate way, resulting in more accurate criminal verdicts.

I think the point has now been sufficiently made that radical eliminativism should be seen as an option of ‘last resort’, at least from the perspective of social institutions such as the criminal justice system. Interestingly, Stephen Stich has argued that there are no clear criteria concerning when an ontologically radical project of eliminativism is justified, and when a less ambitious ontologically conservative project of reduction will suffice. He concludes that since there is “...no easy measure of how ‘deeply and fundamentally different’ a pair of posits are, the conclusion we reach is bound to be a judgment call.”⁴¹ If this is true, our judgment with regard to the role of CSP in the criminal law might be justifiably influenced by the practical concern of maintaining a working criminal justice system. That is, if it is a close call – if it isn’t clear that the commonsense concepts we are using to generate verdicts fail to refer to real psychological phenomena – we might wish to err on the side of reduction.

With this in mind, let’s now explore some of the arguments offered in support of radical eliminativism, specifically looking to see if such arguments provide evidence that CSP is hopelessly false.

⁴¹ Stich, *Deconstructing the Mind*, 94.

A. Arguments for Radical Eliminativism

In Paul and Patricia Churchland's recent book *On the Contrary*, Churchland offers three general arguments regarding why CSP must be radically eliminated as a theory of behavior.⁴² First, Churchland claims that CSP has not 'progressed' – it has not “shown the expansion and developmental fertility one expects from a true theory.”⁴³ Second, Churchland argues that CSP shows no signs of being smoothly integrable with “...the emerging synthesis of the several physical, chemical, biological, physiological, and neurocomputational sciences.”⁴⁴ These sciences are currently merging to provide a holistic description of the world, according to Churchland, and if CSP is not compatible with this description, so much the worse for CSP. Third, Churchland argues that CSP “...fails utterly to explain a considerable variety of central psychological phenomena” like mental illness, memory, intelligence differences, and the different forms of learning.⁴⁵ He claims a ‘true theory’ would not have such explanatory gaps.

These general concerns regarding the accuracy of CSP and its compatibility with scientific psychology are referred to widely in the eliminativist literature, and have also been expressed by those specifically worried about the role CSP plays in the criminal law. In a fairly recent University of Pennsylvania law review article, Andrew Lelling uses these and other arguments in an attempt to raise the alarm about the role CSP plays in the criminal law.⁴⁶ Lelling claims that if the law continues to use CSP to generate verdicts and to ignore new research in cognitive science and neuroscience, there could be “...misapplications of blame – the law would be guided by a morality based on a

⁴² Churchland Paul and Churchland Patricia Smith, *On the Contrary: Critical Essays, 1987–1997* (Cambridge, MA: MIT Press, 1999).

⁴³ *Ibid.*, 7.

⁴⁴ *Ibid.*

⁴⁵ *Ibid.*

⁴⁶ Lelling Andrew E., ‘Eliminative Materialism, Neuroscience and the Criminal Law’, *University of Pennsylvania Law Review* 141, no. 1471 (1992–1993), 1471–1564.

faulty view of behavior, leading to punishment of persons not necessarily deserving of retribution.”⁴⁷

Let’s review each of Churchland’s arguments in turn.

Lelling shares Churchland’s first worry about the ‘progression’ of CSP, arguing that “...any conceptual system that has remained fundamentally intact for thousands of years must be at least suspected of obsolescence...”⁴⁸ In support of his claim Lelling notes other areas of knowledge where we have revised various concepts due to the emergence of new scientific evidence, such as our notions “...that the heavens constituted a huge, slowly turning sphere, with Polaris as its axis.”⁴⁹

Of course, Lelling is right when he says that we now subscribe to an astronomy ‘based upon a profoundly different framework’ due to scientific advancement. But note that we still use many of the same concepts we used 500 years ago: the concepts ‘stars’, ‘planets’, and ‘axis’ for example. Our beliefs about the referents of these concepts have changed significantly; but we haven’t thrown out the old astrological taxonomy wholesale.

Scientific research into how our brains work is a relatively new phenomenon due to our previous lack of scientific tools for performing such research. Our commonsense concepts about mental states have remained relatively unchallenged before the past few hundred years because we didn’t have a science yet developed to provide new evidence of what is going on in the human mind. Hence, psychology is a conceptual system that has remained fundamentally intact until the 20th century.⁵⁰

⁴⁷ Ibid.

⁴⁸ Ibid., 1506.

⁴⁹ Ibid.

⁵⁰ Note that while CSP has remained ‘fundamentally’ intact our beliefs about the referents of its concepts have changed to some extent. For example, after Freud’s work became well-known it became common to refer to ‘unconscious’ beliefs and desires. In addition, over the past 50 years our general knowledge about what happens in the brain when one is insane, intoxicated or mentally retarded has become much more sophisticated. A better understanding of these phenomena has led to judges and juries to better use the concepts when generating verdicts.

But the relatively late development of cognitive science and neuroscience does not comment in any way upon the veracity of CSP concepts. Just because the progression of science allowed commonsense astrological concepts to be revised before commonsense psychological concepts doesn't mean CSP concepts are not revisable. The only way to judge the veracity of CSP concepts (and the possibility of their reduction) is to look at their referents from the perspective of scientific psychology, and not until these sciences become more advanced will we discover the extent to which our beliefs about mental phenomena accurately describe mental phenomena.⁵¹

Churchland also argues that CSP shows no signs of being smoothly integrable with the 'emerging synthesis' of chemistry, biology, physics, and neuroscience. But standing alone this claim begs the question regarding whether CSP can be reduced to a scientific theory of psychology. The emerging synthesis itself does not prove commonsense psychological concepts to be false: it must be shown that CSP is in fundamental conflict with the way these sciences describe the world. If it turns out that CSP can indeed be reduced to a scientific vocabulary, then its place within this synthesis will be clear.

Finally, Churchland argues that CSP fails to explain certain psychological phenomena such as mental illness and memory, and claims a 'true theory' would not have such explanatory gaps.⁵² But it hardly follows from the fact that a theory is incomplete that the claims the theory does make are false. Only evidence

⁵¹ That Lelling evaluates the 'progress' of CSP from this perspective just highlights how he has the wrong end of stick when he claims that "...if folk psychology represents a verifiable science, there should have been at least some discoveries during the past two millennia marking its progress, as there were with every other empirical science". Lelling, "Eliminative Materialism, Neuroscience and the Criminal Law", 1508. No one claims that CSP is an empirical science in this sense. CSP is not a scientific theory; it is a commonsense theory, which by definition means it is a theory used by average 'folk' without use of the tools of science.

⁵² Churchland and Churchland, *On the Contrary: Critical Essays*, 1987–1997.

that commonsense concepts fundamentally fail to refer to real phenomena should convince us that the theory is false.⁵³

It seems that Churchland's general concerns about the nature of CSP – standing alone without the support of specific claims of how CSP falsely describes real psychological phenomena – have failed to provide us with evidence that CSP is false, and thus have failed to support the cause of radical eliminativism. However, there are many more specific claims about the veracity of CSP that must be reviewed before we can determine whether we are justified in continuing to use CSP for the purposes of the criminal law. The most important and damaging arguments are helpfully summarized in Andrew Lelling's law review article. Below I have clumped the arguments discussed into four categories in order to best review their merits.

1. Concern about the formulation of mental contents as propositional attitudes

CSP can be interpreted as assigning meaning or content to mental states as discrete bits of information having a sentence-like structure something like the structure of the language we use to describe them. That is, if 'Fred believes it is raining' or 'Fred sees the umbrella', one way of characterizing Fred's

⁵³ Regardless of its 'incompleteness', many would argue that our default position should be an assumption that the theory of CSP is true, primarily because CSP appears to be so successful as a theory of human behavior. The truth of CSP seems to be the best explanation for the success we enjoy in interpreting and predicting each other's behavior. Further, the difficulty autistics face with social relations may provide some insight into a world where we didn't have the tools of CSP at our disposal.

For more specific arguments regarding the veracity of CSP, see Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. While Fodor admits that the predictive power of CSP isn't 100% accurate, he argues that the ceteris paribus phrases necessary for CSP generalizations are similarly necessary in all of what Fodor calls the 'special sciences'; sciences such as psychology and geography that quantify over different types of entities – 'natural' or 'real' kinds. Fodor argues that the special sciences support counterfactuals even though they have exceptions, because in the special sciences we can't enumerate all the conditions under which the generalizations will hold using the vocabulary of a special science. Fodor, "Special Sciences (or the Disunity of Science as a Working Hypothesis)."

mental states is to see them as a propositional attitude, where Fred has an attitude (a belief or desire) about something specific in the world, such as the rain or the umbrella. Thus Fred's belief that it is raining can be understood as the 'attitude' of belief toward the proposition that it is raining.

It follows that Fred may have beliefs about any number of propositions, which may or may not be true. Thus, characterizing mental states as propositional attitudes has the bonus of allowing us to think about mental states – and thus, to some extent 'thinking' – as a productive process, where attitudes may be paired with an enormous amount of propositions to allow a thinker to think original thoughts.

Lelling worries that scientific psychology has shown that this way of characterizing mental states "...is logically incoherent; the point is not that neuroscience *won't* vindicate beliefs and desire, but the stronger assertion that it *can't*, because of inherent difficulties with the concept itself."⁵⁴ More specifically, Lelling argues that "...There is no reason to suppose that language accurately relates the nature of a brain state. When we are asked what we believe, we internally construct a sentence to be spoken moments later."⁵⁵ Lelling then goes on to claim that such self-reports do not reveal anything 'psychologically important underlying beliefs' but instead reveal the way in which we are forced to describe complex internal states using the tools of language (words). His argument for this conclusion has to do with our hesitation to attribute beliefs to non-human animals such as frogs. Lelling thinks frogs don't have propositional attitudes. "But as we progress on the evolutionary scale, why should it suddenly be obvious that human beings enjoy and employ propositional attitudes?" As the neurology of thinking becomes more complex, it is just easier to 'stick' propositional attitude labels on people, "...even though they are, at best, an extremely distorted description of neurophysiological processes and thus psychology."⁵⁶

⁵⁴ Lelling, 'Eliminative Materialism, Neuroscience and the Criminal Law', 1498.

⁵⁵ *Ibid.*, 1500.

⁵⁶ *Ibid.*

Many philosophers of mind, however, believe that CSP isn't committed to the view that all mental content is 'sentence-sized' or easily characterized as a propositional attitude. It has been proposed that some mental states have non-propositional or 'non-conceptual' content, such as perceptual or emotive mental states.⁵⁷ For example, perceptual states such as visual states may refer to a whole host of information about a particular visual field, and may not be easily formulated as 'belief that there is a book, table, person... there.' Commonsense attributions such as 'Mom saw the state of your room' or 'I know that church is beautiful on the inside' appear to refer to the large quantity of information a single visual perceptual experience can carry. Similarly, Daniel Dennett's famous example of a man who has 'a thing about redheads' is another illustration of a commonsense psychological description of a mental state that is not easily formatted as a propositional attitude.⁵⁸

In short, Lelling's worry about the 'discrete' sentential format of propositional attitudes seems ill conceived. It will not be surprising if the concepts we use to refer to mental states are somewhat clumsy heuristics for complex neurophysiological happenings. Many commonsense concepts are, such as the concepts 'cancer' and 'run' or 'throw'. Just because commonsense psychological concepts use a discrete system of symbols, language, to refer to complicated physical events does not mean that such concepts don't refer to something real in the world. In the end, Lelling doesn't provide an argument that something essential to the commonsense concepts gets psychology wrong such that they don't refer, and thus his concern about the way CSP formulates mental content seems unfounded.

2. Worries about the language of thought

Lelling is also worried about a specific theory of how the mind works based upon a view of mental states as sentential.⁵⁹

⁵⁷ See Dretske and Peacock's chapters in York Gunther (ed.), *Essays on Nonconceptual Content* (Cambridge, MA: MIT Press, 2002).

⁵⁸ Dennett Daniel, *Things About Things* (1998).

⁵⁹ Lelling, 'Eliminative Materialism, Neuroscience and the Criminal Law', 1504.

Jerry Fodor, the realist mentioned briefly above, postulates that thinking is the processing of discrete mental symbols (which may represent beliefs and desires) which can be manipulated in such a way to produce an infinite number of possible thoughts.⁶⁰ According to Fodor, CSP is the capacity to recognize that such symbols are the cause of other people's behavior, and one uses CSP to postulate the particular mental symbols that have been or will be tokened in someone else's head as a means for understanding and predicting their behavior.

Lelling claims that Fodor's theory—the language of thought hypothesis, or LOT—must be wrong, because "...Under sententialism, each person, in order to believe the same thing as someone else, must have the exact same neuronal components arranged in relation to each other in exactly the same way."⁶¹ This simply is not true. Fodor does not make specific claims about the way his 'tokens' are realized at the neuronal level. It may be that two symbols could have the same syntactic/causal properties within a system, but still be realized in different neurons or different neuronal patterns.

Regardless of whether LOT is true, however, one must keep in mind that it is just one theory compatible with CSP about how mental content is represented and processed. Even if it is wrong, some other theory that makes use of commonsense psychological concepts of mental states may turn out to be right. Thus Lelling's concerns about LOT need not translate into concerns about the veracity of CSP.

3. Truth conditions and wide vs. narrow content

Lelling is also concerned about the fact that propositions are generally thought to have truth conditions (conditions under which they are true). That is, Fred's thought that it is raining may or may not be true, depending on whether or not it is raining. Lelling attempts to flesh out his worry by citing philosopher Hillary Putnam's well known H₂O/XYZ argument, an

⁶⁰ Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*.

⁶¹ Lelling, 'Eliminative Materialism, Neuroscience and the Criminal Law', 1504.

argument whose aim is to convince readers that assigning content to mental states may in some cases depend upon the environment a thinker is in.⁶² The commonsense concept 'water' may mean H₂O in our world, but in another world, that colorless tasteless liquid we drink may be XYZ. Similarly, my mental state about 'the population of New York City' would seem to depend in some sense on the actual population of New York City.⁶³ Strangely, Lelling concludes from this "...the way folk psychologists want to individuate mental states and objects is by how they refer to things in the world, but they cannot be in your head because they are in world."⁶⁴

Putnam's argument is meant to show that that mental content should be formulated (or that mental contents be 'individuated') in a particular way – 'widely', or with reference to the thinker's environment. For this to be a problem for CSP, one must assume that scientific psychology has shown that brain states are individuated narrowly, or without reference to anything outside the thinker's head. In this case CSP would appear to be describing mental content inaccurately.

But there is an easy response to this sort of argument. Lelling provides no argument that scientific psychology is committed to internalism, or narrow content. Any claim that scientific psychology was committed to internalism would be hotly disputed by many who argue that we should expect scientific psychology to vindicate an externalist theory of mental content. Without evidence that a commonsense psychological concept is in fundamental conflict with a scientific description of mental states, we again have no evidence that the commonsense concept doesn't refer to some real mental phenomena in the world.

4. Identification of CSP concepts with brain states

Finally, Lelling has concerns about specific attempts to show how CSP concepts might be found to be compatible with

⁶² Putnam Hilary, *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975).

⁶³ Lelling, 'Eliminative Materialism, Neuroscience and the Criminal Law', 1502.

⁶⁴ *Ibid.*, 1503.

scientific psychological evidence. Identity theorists postulate that mental states actually are just physical states in the brain. 'Type' identity theories claim that the mental events postulated by CSP will be found to be identical with physical events. Events are thought to be structured particulars consisting of objects, properties and times.⁶⁵ Type identity is thus a direct identification of properties: for a mental and a physical event to be the same event it must be an instantiation of the very same property by one object at one time.

This is a very strong claim, whereby a mental event – a belief that 'P' – is generally identified with a particular mental state. Every time one attributed to another a particular mental state, they would also be attributing a specific physical neuronal event to that person, just as every time one called something water (on Earth) they would be referring to a specific molecular entity (H₂O).

Lelling notes that it would be "...a coincidence of some magnitude for our relatively crude CSP to correspond, state for state, to the workings of the most complex structure ever encountered by humanity."⁶⁶ If what he means is that it would be surprising if our rough and ready commonsense notions of psychology sliced up brain states at the neuronal level, he's right. It seems unlikely that when we attribute to both Wilma and Fred the belief that it is raining, we are attributing to them the exact same neuronal state.

But type identity theory is not the only option for postulating relationship between mental states and brain states. Another such theory is functionalism. Functionalists are motivated by the observation that mental states such as the state of being in pain seem to be 'multi-realizable' – that is, realized in very different sorts of physical events. Not only does it seem possible that Fred and Wilma physically instantiate pain differently, but it appears that animals as far removed

⁶⁵ Kim Jaegwon, 'Events as Property Exemplification', in *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press, 1993), David Papineau, *Philosophical Naturalism* (Oxford: Blackwell, 1993).

⁶⁶ Lelling, 'Eliminative Materialism, Neuroscience and the Criminal Law', 1511.

from us as octopi and ducks may experience pain. Functionalism thus identifies mental states with functional or causal roles, and not directly with brain states. Pain, and other mental states, are to be individuated by their characteristic patterns of relations to their inputs (generally, perception), to their outputs (generally considered action or behavior) and to other mental states. Under functionalism, then, the mind is a physical system or device where external or environmental states induce changes in the system's internal states, which cause other internal changes, eventually leading to a determination of the system's overt behavior.

Lelling is worried about functionalism partly because it "...would warrant extending legal personhood to robots and other systems capable of the same functional representations as humans."⁶⁷ While it is unclear whether or not this is true, Lelling's concern doesn't articulate a problem with functionalism per se. It articulates a potentially undesirable outcome of the theory. What we are concerned with is whether or not it is possible that CSP terms will be able to accommodate evidence coming from science about how the brain works. Functionalism is one way in which CSP mental state concepts may truly characterize physical systems. And if functionalism does turn out to be a true characterization, the possibility of applying the criminal law to robots that have intent to kill is an issue completely separate to the issue of the veracity of the theory.

Lelling is also concerned that functionalism gives too vague an account of mental states to be satisfying, because they do not refer to the underlying neuronal identity of the states. But it isn't clear that one needs to reduce mental states to neuronal states in order to say something true about the physical or causal identity of such states. If functionalism can accurately explain and predict mental states and their relationships to inputs and outputs, it wouldn't seem to matter that this prediction was not based on an understanding of how particular neurons instantiate these relationships.

In addition, functionalists can make two different kinds of identity claims between mental states and functional roles. An

⁶⁷ Ibid., 1513.

identification can be made between a mental property and a second-order functional property – the property of being a state with a causal role such that it mediates between such and such inputs and such and such outputs.⁶⁸ However, an identification can also be made with the first-order property of whatever physical entity *realizes* the second-order causal role.⁶⁹ In this case, being in pain would be identified as a disjunction of whatever states it is that physically occupy such and such causal role.

If one is attempting to make generalizations across human and octopi pain, the first-order identity of a mental state like pain may be very unruly. However, for the purposes of our project we are only worried about human mental states (and possibly robots with human-like states). This sort of functionalism, or a version of functionalism which doesn't require a tight reduction to neuronal states, may both be accurate and specific enough for the purposes of the law.

Finally, functionalism isn't the only theory other than type identity that attempts to describe the relationship between mental and physical states. Token identity theory is also an option. Token identity claims that an event is a particular that can be described both in a mental and in a physical way. However, for a mental and physical event to be the same event its physical and mental properties do not need to be strictly identified. The mental and physical properties are seen as 'arising' from or tacked onto the same event, just as every baseball has the properties of being both shaped and colored. Thus, the mental properties do not need to be 'reduced' to the physical.

In this formulation, token identity is quite a weak claim because it doesn't say anything about the relationship between the mental and physical properties of an event. However, Donald Davidson has strengthened the position by arguing that mental properties *supervene* on physical properties.⁷⁰ A set of

⁶⁸ Putnam Hilary, *Representation and Reality* (Cambridge, MA: MIT Press, 1988).

⁶⁹ Lewis David, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (1972): 249–258.

⁷⁰ Davidson Donald, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

properties or facts M supervenes on a set of properties or facts P if and only if there can be no changes or differences in M without there being changes or differences in P. This means that mental states are in some sense dependent on physical states. “Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.”⁷¹

One example of supervenience might be the relationship between an object’s physical and aesthetic properties. That is, with regard to a painting, the property of being beautiful supervenes upon the physical properties of the paint and canvas.

Token identity plus supervenience thus offers us another possible relationship between the mental and the physical that does not require strict identity theory. And given that token identity and functionalism remain live options, it seems that Lelling is too quick to declare that we will never make sense of the relationship between commonsense sense mental state concepts and the brain states they appear to refer to.

B. Conclusions about Radical Eliminativism

Lelling’s worries about particular formulations of commonsense psychological theory and their compatibility with scientific psychology seem at best, premature, and at worse, misguided. Neither Churchland’s general concerns about the nature of CSP nor more specific claims about the veracity of CSP have provided us with evidence that CSP concepts fundamentally fail to refer to real mental phenomena. Thus it seems we may still hold out hope that the CSP concepts used in the criminal law will be compatible with emerging scientific theories of psychology.

V. GENERAL CONCLUSIONS

Eliminativism about CSP must be recognized for the radical project that it is: a project whose aim is to throw out a sub-set

⁷¹ Lelling, ‘Eliminative Materialism, Neuroscience and the Criminal Law,’ 98.

of commonsense concepts vital to the everyday task of understanding and predicting behavior. For such a project to be justified, commonsense psychological concepts would have to fail to refer to something in the world, and thus have no chance to be reinterpreted or redefined to accommodate a scientific understanding of their referents or content. Above we reviewed various arguments in support of the radical eliminativist position, and in each case we discovered there is no viable evidence from scientific psychology that commonsense psychological concepts fail to refer to mental entities.

It is good news for the current criminal justice system that the possibility CSP could be reduced to scientific psychology remains open. If the CSP concepts used in the criminal law can be reduced to some version of scientific psychology, this science may be used to improving our understanding of the referents of the CSP concepts used in the criminal law. As a result, we may begin to apply these concepts in a more accurate way, resulting in more accurate criminal verdicts.⁷²

Rockefeller Fellow in Law and Public Policy
Dartmouth College
Thornton Hall
Hanover, NH 03755
USA

⁷² Many thanks to David Papineau and Michael S. Moore for invaluable comments on earlier drafts of this article. I am also indebted to Walter Sinnott-Armstrong, Susan Brison and the Rockefeller Center for their support during my fellowship at Dartmouth College.