CDD: 128.2

A NONREDUCTIONIST'S SOLUTION TO KIM'S EXPLANATORY EXCLUSION PROBLEM

JEELOO LIU

Department of Philosophy
SUNY Geneseo
1 College Circle
GENESEO, NY 14454
USA

Liu@geneseo.edu

Abstract: *In this paper I shall give a close examination of Kim's argument for the exclusion problem. Kim thinks that the only solution to the exclusion problem is reductionism. I shall propose another solution that does not require that mental properties be reducible to physical properties, or that psychology be reducible to neuroscience. My paper is divided into three parts. In the first part I shall explain how the explanatory exclusion problem is generated. In the second part I shall examine Kim's argument for the exclusion of psychological explanation, namely, his supervenience argument. And finally in the third part, I shall offer my solution to this problem. In a nutshell, my solution will be that of mental/physical properties identity defined locally, holistically. My analysis is based on Davidson's mental/physical event identity. An individual's mental event is nothing but the physical event of the individual's brain. But since the properties we use to specify a mental event and the properties we use to specify its corresponding a physical event must be defined through the whole system in which the event in question takes place, these properties cannot be the properties dealt with in either psychology or neuroscience. I shall explain how this analysis of identity does not commit one to mental/physical properties reductionism, or intertheoretic reductionism between psychology and neuroscience.*

Key-words: *Jaegwon Kim; Donald Davidson; Nonreductive physicalism; mental causation; holism.*

In numerous papers Jaegwon Kim argues that nonreductive materialists (i.e., those philosophers who believe that there are no irreducible non-physical objects in the universe, and yet there are

© *Manuscrito* 2001. Pubished by the Center for Logic, Epistemology and History of Science (CLE/UNICAMP), State University of Campinas, P.O. Box 6133, 13081-970 Campinas, SP, Brazil.

irreducible psychological properties which are indispensable in intentional psychological explanations) face two problems. One is that intentional mental properties are not causally relevant; the other is that explanations appealing to these properties are excluded by explanations appealing to physical, in particular, microphysical, properties.[1] The first problem can be called the problem of epiphenomenalism. The second problem is what Kim calls the problem of explanatory/causal exclusion. Epiphenomenalism is not just a problem for nonreductive materialists, but a problem for psychology (as a science of psychological explanation), and for anyone who believes that our thought is the cause of our action. Kim argues that the exclusion problem, on the other hand, is especially a problem for nonreductive materialists. Nonreductive materialists typically buy into *the principle of the causal closure of the physical domain*, according to which "any physical event that has a cause at time *t* has a physical cause at *t*..... [I]f we trace the causal ancestry of a physical event, we need never go outside the physical domain." (Kim (1989b), p. 280). One cannot reject this principle unless one is willing to go back to Cartesian dualism. But the problem is, if all causation takes place at the physical level, then there is no causal work left for mental kinds. If mental kinds do not play any causal role, then intentional psychological explanations that appeal to mental properties are not warranted. Therefore, Kim concludes, nonreductive materialism is an untenable position.

Some nonreductionists (Davidson, Dretske, Lepore & Loewer, Jackson & Pettit, for example) have tried to argue for the causal relevancy of mental properties from an explanatory point of view; i.e., they argue that mental properties are causally relevant because they play a significant role in psychological explanations of behavior. In other

[1]For Kim's discussions on the explanatory exclusion problem, see Kim (1987a), (1989a), (1989b), (1990), (1993b), (1993c), (1993d), (1995), (1998a), (1998b).

words, they claim that mental properties can be causally *relevant* without being causally *efficacious*. Kim, however, argues that this approach does not work. He thinks that a causal explanation is one that gives the causal history of an event. If we want to say, for example, that our desire for water causes us to drink water, then our desire must indeed *cause* us to drink water if this explanation is correct. But we also say that the neurophysiological state of our having that desire is the cause of the neurophysiological state of our water-drinking, since all the causal work is done at the neurophysiological level. There is only one causal history. Hence, only one explanation can be giving us the real causal story. Since nonreductionists do not deny that causation takes place at the neurophysiological level, mental explanation cannot give a different causal history. Kim argues that mental explanation is thus excluded by neurophysiological explanation. If Kim is right, then the problem of epiphenomenalism would ensue from the problem of explanatory exclusion. Mental properties would not be causally relevant exactly because they are not explanatorily relevant. Anyone who does not want to endorse epiphenomenalism would thus have to take the problem of explanatory/causal exclusion seriously.

In this paper I shall give a close examination of Kim's argument for the exclusion problem. Kim thinks that the only solution to the exclusion problem is reductionism.[2] I shall propose another solution that does not require that mental properties be reducible to physical properties, or that psychology be reducible to neuroscience. My paper is divided into three parts. In the first part I shall explain how the explanatory exclusion problem is generated. In the second part I shall examine Kim's argument for the exclusion of psychological

[2]Kim says, "At this juncture it seems highly plausible that the only solution to the exclusion problem and the problem of the physical causal closure lies in some form of reductionism that will permit us to discard, or at least moderate, the claim that mental properties are distinct from their underlying physical properties." (1993b), p. 356.

explanation, namely, his *supervenience* argument. And finally in the third part, I shall offer my solution to this problem. In a nutshell, my solution will be that of mental/physical properties identity defined *locally*, *holistically*. My analysis is based on Davidson's mental/physical event identity. An individual's mental event is nothing but the physical event of the individual's brain. But since the properties we use to specify a mental event and the properties we use to specify its corresponding physical event must be defined through the whole system in which the event in question takes place, these properties cannot be the properties dealt with in either psychology or neuroscience. I shall explain how this analysis of identity does not commit one to mental/physical properties reductionism, or intertheoretic reductionism between psychology and neuroscience.

1. THE GENERATION OF THE EXPLANATORY EXCLUSION PROBLEM

The problem of explanatory exclusion is based on a principle of explanatory exclusion, namely:

[EE] Distinct (causal) explanations of a single explanandum exclude one another, in spite of the fact that their explanatory premises are mutually consistent. (Kim (1990), p. 123)

Taken by itself, [EE] seems too strong. And indeed it has aroused some suspicion that Kim is creating a pseudo-problem.[3] However, [EE] is generated by several assumptions that Kim makes, and these assumptions all seem to have some plausibility. In this section we shall take a close look at each of these assumptions.

[3]Davidson and Burge both seem to think that this is not a real problem as long as we recognize the difference between explanation and causation. See further details later.

1. Causal Realism

[CR] Causality itself is an objective feature of reality.[4]

[CR] is posed against a *Humean* analysis of causation. [CR] can also be viewed as a thesis of the *extensionalist's* position on causal relations, according to which causality is a relation among events themselves. Davidson, for example, is an advocate of causal extensionalism. I shall leave this assumption as uncontroversial in the context of the debate.

2. Causal Exclusion

[CE] For any single event E, there cannot be two causes C and C^*, each of which is a sufficient cause of E, while they are distinct and independent of each other.

Kim considers five possible cases when we identify both C and C^* as the causes of E:

Case 1. Identity: $C = C^*$. In this case the two causes are not competing with each other since there is only one cause here.

Case 2. Reduction or supervenience: C is distinct from C^*, but is reducible to, or supervenient on C^*. According to Kim, in this case C is not an independent cause from C^* and thus the causal competition does not exist.

Case 3. Neither C nor C^* is a sufficient cause of E, though each is an indispensable component of a sufficient cause. This again is an acceptable situation since neither cause is taken to be sufficient by itself.

Case 4. C and C^* are different links in the same causal chain leading to the occurrence of E. Again, Kim argues that this case does

[4] Kim (1987a), p. 229. Kim takes this position to be countering Hume's view on causality.

not constitute a violation of [CE] since one of the two causes is causally dependent on the other cause.

Case 5. C and C^* are distinct and each a sufficient cause of E. This is the case that is barred by [CE]. According to Kim, this is a case of *overdetermination*. "E would have occurred even if either C or C^* had not occurred, or had not caused it; the other would have been sufficient to bring it about." (Kim (1989a), pp. 250-253)

It is not clear whether there can really be cases of overdetermination. And even if there are, we don't want to view the competition between mental causation and neurophysiological causation to be a case of overdetermination. We also don't want to view the connection between the two causes to be an example of Case 3 or Case 4. The only remaining options are Case 1 or Case 2; namely, identity or reduction. Kim thus argues that if the mind causes, then it is either identical to the brain or reducible to the brain and the two analyses are compatible with each other.

[CE] is not contested by most nonreductive materialists. The question is whether it can be extended to [EE]. Both Burge and Davidson think that causation is one thing while causal explanation is another, and there is no warranty in extending the exclusion of two causes to the exclusion of two explanations. Davidson argues that Kim is committing a slippage in his argument: "It is only if we confuse causal relations, which holds only between particulars, with causal explanations, which, so far as they are 'sufficient' must deal with laws, and so with types of events, that we would be tempted to accept the principle of 'causal-explanatory exclusion'." (Davidson (1993), p. 16) Kim, on the other hand, thinks that causal exclusion and explanatory exclusion are one and the same problem. He and Davidson or Burge simply differ on the metaphysics of explanation. For Kim, a basic metaphysical assumption concerning explanation is what he calls the explanatory realism.

3. Explanatory Realism

[ER] C is an explanans for E *in virtue of* the fact that c bears to e some determinate objective relation R (Kim (1989a), p. 226).

Davidson, for one, does not endorse [ER]. Davidson separates the issue of causation and the issue of explanation. He treats the first as a metaphysical issue and the second as an epistemological issue. Davidson stresses that accurate causal explanations are simply adequate descriptions of the causal history of events, and different causal explanations merely *re*describe events. He says: "Redescribing an event cannot change what it causes, or change the event's causal efficacy." (Davidson (1993), p. 7) Let R_1 stand for the causal relation depicted by a psychological explanation; let R_2 stand for the causal relation depicted by a neurophysiological explanation. If psychological explanation merely gives one description while neurophysiological explanation gives another, then R_1 and R_2 are not "determinate objective relations." R_1 and R_2 are merely different relations established through linguistic variation and theory construction.

Burge also does not seem to endorse [ER]. Burge views explanations as giving descriptions of the *pattern* exemplified in singular events. Burge argues, "The existence of a closed system reflects a pattern of causal relations and of causal explanation that needs no supplementation from the outside. There are no gaps. *It does not follow from this that such a system excludes or overrides causal relations or causal explanation in terms of properties from outside the system.*" (Burge (1993), p. 102) In other words, on the *pattern* view, there can be one set of causal relations established within the psychological system and another set established within the neurophysiological system. The two sets do not compete with each other since they only reflect psychological patterns and neurophysiological patterns. Different patterns of causal relations are defined through each discipline's theoretical and methodological

constraints. Therefore, the set of psychologically established causal relations does not compete with the set of neurophysiologically established causal relations that apply to the same objects.

Both Davidson's and Burge's counter claims can be met with by another implicit assumption Kim makes concerning the nature of explanation. In contrast to Burge's *pattern* view of explanations, Kim argues that explanations are "represented by singular causal propositions." (Kim (1987), p. 230) We can formulate this assumption as Kim's propositional view of explanation.

4. The Propositional View of Explanation

[PE] 'Having' an explanation is simply a matter of knowing a certain proposition to be true. (Kim (1987), p. 231)

The main difference between the pattern view and the propositional view of explanation, in Kim's words, is that "gaining a new explanation, on the pattern view but not on the propositional view, does not necessarily involve acquiring new information about facts of the world." (Kim (1987a), p. 231) That is to say, the pattern view would allow different explanations of the same "fact" while the propositional view demands one unique explanation of the fact. Since for every event there is a "*unique and determinate causal history*", under Kim's [ER], for every event there should only be one unique singular causal proposition that is true. That is to say, if there is only one event causing, then there is only one true propositional causal explanation.

How do we reconcile the two opposing views on the nature of explanation? On the one hand, both Davidson and Burge seem to be right in pointing out the epistemological and theory-laden nature of explanation. Explanation can be given in different fields and each field has its own taxonomy and its own laws. Explanations represent our understanding of the external world. An expert from a different field

can give us an explanation that carries with it significant information. If we have two explanations that have different cognitive contents, there is no reason why we must claim that only one of them is true. There is also no reason why we must say that both of them must be false because neither gives the uniquely true proposition.

But on the other hand, Kim could argue that even if those non-causal, structural dependence or *pattern* explanations are of epistemic value, what he is interested in is only causal explanation. In other words, Kim could grant that non-causal explanations are aimed to *redescribe* events, and different theories have different vocabulary and laws to offer different explanations. But when it comes to causal explanation, our interest is only in finding out the true causal story. Kim does not need to embrace an explanatory monism, the view that only one explanation is the real explanation.[5] Even though he defends [ER] in many articles, he really does not need such a strong premise to argue for the exclusion problem. What he needs is a more restricted principle as follows:

5. Causal Explanatory Principle

[CEP] A causal explanation of event *e* in terms of event *c* ("*e* occurred because *c* caused it") is *correct*, or *true*, just in case *c* did as a matter of objective fact cause *e*. (Kim (1989a), p. 229)

According to [CEP], to give a *true* causal explanation, is to cite a causal relation that actually exists between two events. Any other explanation, however helpful in enhancing our understanding of relationships between events, is not a true *causal* explanation.

[5]It is questionable whether Kim himself really endorses explanatory monism. Marcelo Sabatès, for example, argues that Kim at times defends explanatory pluralism, viz., the view that "there are, in addition to causal explanations, explanations tied to non-causal, structural dependence relations." (Sabatès (1996), p. 94).

Is [CEP] still too strong? Maybe so. However, it is a reasonable principle as long as Kim also grants other non-causal explanations to be legitimate. A true causal explanation gives a true proposition of the causal relationship. It does not have to be covered under any causal law, and thus it is not restricted to the laws and vocabulary of any discipline. It may not even be available at the present stage of our scientific knowledge. What Kim demands is simply a single proposition that cites the actual cause of a given event. *This* proposition is what he deems the genuine causal explanation. There may be other explanations that make the connections between events more intelligible to us. They are thus of epistemic value. But on the metaphysical level there can only be one causal connection, and what Kim's [CEP] says is that there can only be one causal explanation that really describes this causal connection. Kim says, "To be in need of an explanation is to be in an epistemically incomplete and imperfect state, and to gain an explanation is to improve one's epistemic situation; it represents an epistemic gain. However, knowledge must involve the real world: to know that *p* requires the truth of *p*, and to have a causal explanation of an event requires that the event specified as its cause be, in reality, a cause of that event." (Kim (1989a), p. 256) In other words, Kim is interested in a particular kind of explanation, one that gives us the true depiction of the real causal history. Even if we can argue that other kinds of explanations are epistemically useful, we cannot deny that what he demands is a reasonable expectation of the function of causal explanation.

If we are to give a psychological explanation of action *e*,[6] then, under [CEP], what our explanation says is that the mental event *c* characterized by the mental property or mental kind *C* did indeed cause

[6]I use a more neutral term 'action' here which includes both 'behavior' and 'physical movement'. Dretske makes a distinction between action and behavior as a way out of Kim's explanatory exclusion problem, but I don't think that really avoids the problem.

e. But at the same time, a neurophysiological explanation that cites neural event *n* gives a *complete and sufficient* causal explanation of action *e*. According to [CE], there cannot be two independent and complete causal relations R_1 and R_2 in any given causation. We therefore need to, asides from explaining the connection between the mental event *c* and the neurological event *n*, explain the connection between R_1 and R_2. This would require us to explain the connection between the mental type *C* and the neurological type *N*. If *C* and *N* are not coextensive, or one is not reducible to the other, then we are saying that there are two independent and complete causal relations R_1 and R_2 in the case of action causation. This is not acceptable. So one relation has to go. Since R_2, the causal relation between the neural event type *N* and the action type *E*, is an objective fact, it seems that it is R_1, or the psychological explanation which specifies R_1, that is to be excluded.[7] This leads us to the problem of explanatory exclusion. That is to say, if any other independent cause is excluded by this causal connection, then any other causal explanation is thereby excluded by the explanation that cites this causal connection. Since this problem is not separable from Kim's causal exclusion principle, Kim usually combines the two terms and calls it the explanatory/causal exclusion.

We can further add another assumption that Kim accepts, which leads one from causal exclusion directly to epiphenomenalism:

6. Alexander's *Dictum*

[AD] To be real is to have causal powers.

Kim simply rejects Davidson's and Burge's treating events/causation as a metaphysical issue, while treating properties/explanation as an epistemological issue. He argues, "[T]urning away from metaphysics to

[7] In Kim (1998b), Kim has a further developed argument for the same conclusion. I shall discuss this argument in the next section.

embrace epistemology, or away from causation to embrace explanation, will not dissipate the need for an account of mental causation." (Kim (1998b), p. 67) This account that Kim demands, is not about *whether* mental causation is possible, but about *how* mental causation is possible. Under [CEP], if mental explanations are to be correct or true explanations, then mental properties cited in these explanations must also have real "causal powers". In giving causal explanations, we simply cannot separate causal efficacy from explanatory relevancy. If mental events do not have causal powers, then, they are not explanatorily relevant. Furthermore, by [AD], they are not *real*. In other words, mental realism is not tenable if we cannot resolve the exclusion problem.

In brief, explanatory exclusion and causal exclusion, for Kim, are one and the same thing. Kim argues, "[W]e think of these [causal] explanations as directly tied to the actual causal histories of the events being explained, and their 'correctness' as explanations is determined by the accuracy with which they depict the causal connections as they exist. When two such explanations of one event are on hand, we need to know how they are situated in relation to each other on the causal map of the event; *one thing that cannot be allowed to happen, if our explanations are to be coherent, is that they tell two different stories about the same region of the causal map*" (italics mine) (Kim (1989a), p. 255) What he is pushing for is basically a recognition of the causal exclusion problem. Explanatory exclusion is derived from the causal exclusion principle plus causal realism and causal explanatory principle. Causal realism is not in contention between Kim and nonreductive materialists. And I have argued that his causal explanatory principle is a plausible principle. But even if we reject his causal explanatory principle and thereby reject his explanatory exclusion principle, we still need to deal with the problem of causal exclusion. Mental properties can still become causally inert or causally irrelevant under the principle of causal exclusion, because they would be competing with the neurophysiological properties appealed to in neurophysiological explanations. Unless we find a way to link mental

properties with neurophysiological properties, we cannot dissipate the tension between the two sets of properties, each competing to be the properties that are causally relevant.

2. KIM'S SUPERVENIENCE ARGUMENT AND KIM'S REDUCTIONIST'S SOLUTION

In Kim's most recent book, *Mind in A Physical World*, he comes up with another argument against the causal efficacy of mental properties. His argument is called the *supervenience* argument, because it is based on a mind-body supervenience thesis that both Kim and the nonreductionists endorse. The following is a (somewhat) simplified version of this argument: (Kim (1998b), pp. 38-46)

1. [The supervenience thesis]: Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t, there is a physical base property P such that the thing has P at t, and necessarily anything with P at a time has M at that time.

2. Suppose that an instance of mental property M causes another mental property M^* to be instantiated.

3. M has a physical supervenience base P, while M^* has a physical supervenience base P^*.

4. But according to the supervenience thesis, M^* is instantiated because its physical supervenience base P^* is instantiated on this occasion.

5. Therefore, as long as P^* occurs, M^* must occur regardless of whether or not an instance of M preceded it.

6. Therefore, P^* alone is the complete and sufficient cause of M^*.

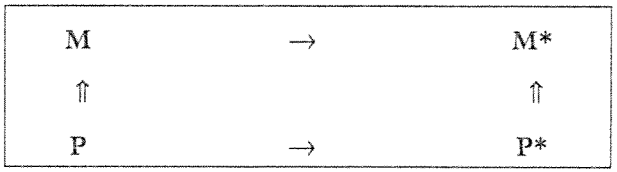
7. For M to be causally related to M^*, M must be a cause to P^*.

8. But P alone is already causally sufficient for P^*.[8]

9. If M also causes P^*, it would either violate the physical causal closure principle or would be a case of overdetermination.

10. Therefore, there is no real causal link between M and M^*.

In this argument, M precedes M^* and P precedes P^*, and thus $M \neq M^*$ and $P \neq P^*$. Distinct mental/physical events have distinct properties. The connection between M and M^*, or P and P^* respectively, is *supposedly* causation. The connection between M and P, or M^* and P^* respectively, is supervenience. Under the supervenience thesis, the supervenience base property necessitates its supervening property. It is accepted by the nonreductionists that there is complete causation between P and P^*. What remains to be established is the causal connection between M and M^*. But since P causes P^* and P^* necessitates M^*, the causal connection between M and M^* becomes superfluous. Using '$\rightarrow$' to represent causation, and '$\Uparrow$' to represent necessitation, we can further simplify the above argument into the following diagram:



[8]Kim's lemma for this step of the argument is twofold: First, if we take causation to be grounded in nomological sufficiency, then we see that since P is sufficient for M and M is sufficient for P^*, P is sufficient for P^*. Second, if we take causation to be grounded in counterfactuals, then if P hadn't occurred, M would not have occurred. And if M had not occurred, P^* would not have occurred, we can conclude that if P had not occurred, P^* would not have either. (Kim (1998b), p. 43)

In the previous section, we have shown that the exclusion of two independent and sufficient causes and the exclusion of two complete explanations appealing to different sets of properties go hand in hand. Now according to this supervenience argument, when there is causal/explanatory exclusion, it is mental properties that must go otiose. The only way to preserve the causal efficacy and the reality of the mental, *if one accepts Kim's reasoning*, is thus to find a way to show that the apparent two causes (mental and physical) are not two separate causes, and the apparent two sets of properties (M and P) are not two independent sets of properties. In other words, either an identity (Case 1) or a reduction relation (Case 2)[9] between the two sets of properties will have to be established in order to solve the exclusion problem.

Kim's own solution to the explanatory/causal exclusion problem is to dissipate the causal competition by a disjunctive identification between mental and physical properties. He calls this theory *multiple-type physicalism* (or MTP for short), by which he means that a mental kind M is *disjunctively identified* with physical kinds P_1, P_2, ..., where P_1, P_2,... are all "realizers" (or, the supervenience bases) of M. (Kim (1993c), p. 364) In Kim's analysis, mental properties are second-order functional properties, "properties specified in terms of their roles as causal intermediaries between sensory inputs and behavioral outputs." (Kim (1998b), p. 19) These causal roles are realized by physical properties since the mental realm has the physical realm as its base. Kim calls this view 'physicalist functionalism.' In such a picture, mental properties can have different physical bases. As Kim puts it, "any base properties with the right causal/nomological relations to other properties can serve as its realizers." (Kim (1998b), p. 21) Different biological species and different cognitive structures can all have different physical bases. But as long as they manifest the same causal roles, they are said to have

[9]Or, the combination of both relations. This combinatory solution is Kim's proposal.

the same mental properties. Mental properties are thus identified *functionally*, and they are identifiable with, and reducible to, a disjunctive set of physical bases, each of which may realize the same mental properties differently. On the basis of this disjunctive identification, Kim claims, mental properties are reducible to physical properties. This is Kim's model of reductionism. And since mental properties are reducible to physical properties, there is no real causal/explanatory competition between the two sets of properties. Mental properties are thus not causally excluded by physical properties.

Kim's reductionism is different from the standard model of intertheoretic reductionism as explicated by Ernest Nagel. According to Nagel's model, for a theory to be reducible to another theory, there must exist universal bridge laws connecting the predicates of the two theories. These bridge laws are expressed in a biconditional form (as, for example, *M* iff *P*), such that for every property of the reduced theory, there would be a nomologically coextensive property of the base theory. By the same token, for every law in the reduced theory, there would be an *a priori* derivation from this law to a law in the base theory. It is against this Nagelian model of reductionism that Fodor provides a celebrated proof that special sciences are not reducible to basic physics.[10] It is also against this Nagelian model of reductionism that Davidson argues that the mental realm is anomalous and thus, irreducible to the lawful physical realm. Now Kim's model of reductionism is not based on the existence of universal bridge laws. In Kim's model, as he puts it, "the reducibility of a property critically depends on its *functionalizability* – whether or not it can be construed as a second-order functional property over properties in the base domain – not on the availability of bridge laws. Bridge laws are neither necessary nor sufficient for reduction." (Kim (1998b), p. 27) That is to say, as long as we can define a mental property – the property of

[10]Fodor, 'Special Sciences,' in Fodor (1983), Chapter 5.

having a pain, using Kim's favorite example – in terms of its sensory inputs (having one's finger pinched) and its behavioral outputs (wincing or saying 'ouch'), we can reduce this mental property to its physical base property. Because the same mental property could be *multiply realized* by different physical properties in different biological species, the reduction is generally not a one-to-one reduction but a one-to-many reduction. We can perhaps summarize the difference between the two models of reductionism in this way: on the Nagelian model, between mental properties and physical properties it is expected to have biconditional laws in the form of *M* iff *P* (where *P* is a single physical or neurophysiological property); on Kim's model, it is acceptable to have *M* iff $\{P_1, P_2, ..., P_n\}$ where $P_1 \neq P_2 \neq \ldots \neq P_n$.

Kim's model of reductionism is an improvement over the standard model. It is not subject to the multiple realizability argument that is typically used against the standard model of reductionism; it is actually based on multiple realizability. Kim's model also changes the focus of debate between reductionists and anti-reductionists. The issue is no longer whether mental properties are multiply realizable or whether mental properties are anomalous, both of which have been used to defend nonreductionism. Kim acknowledges that they are. This issue now is whether mental properties are functionalizable. Kim thinks that the only kind of mental properties that are (possibly) not functionalizable are qualia or what Ned Block calls phenomenal consciousness. (Kim (1999), pp. 9-11) But under Kim's view, if these properties are not functionalizable, then they are not reducible; if they are not reducible, then, by his Exclusion Principle, they are causally inert and irrelevant. Finally, if they are causally inert, then by Alexander's Dictum, they are not real. Therefore, Kim concludes, the only real mental properties are the ones that are reducible to physical properties.

However, what most nonreductionists are arguing against is only intertheoretic reductionism, not ontological reductionism between the

mental and the physical.[11] That is to say, the debate between the reductionists and the nonreductionists is not about the ontology of the mental, but on the explanatory power of general physics. Intertheoretic reduction may not be reliant on the existence of singular bridge laws; it nonetheless has to imply that the reduced theory can be fully and completely explained in terms of the base theory. Thus, even if Kim is right that to establish intertheoretic reductionism, he does not need to demonstrate that there is an exact one-to-one match between psychological kind predicates and physical kind predicates, he still needs to demonstrate that all psychological kind predicates can be fully and completely analyzed in some physical kind predicates. I argue that Kim's argument does not give us this latter demonstration. In his example, the functionalizability of the property of pain is specified in terms of its sensory inputs and behavioral outputs, and it seems possible that the descriptions of sensory inputs and behavioral outputs could be based on the vocabulary of physics (or neurophysiology) alone. Kim calls the states or properties that satisfy this causal specification (of inputs and outputs) "the 'realizers' of the mental property involved." (Kim (1999), p. 10) Thus, once the mental property can be functionalized, it has found its physical realizers. But the success of this example may have given Kim the false hope that all mental properties can be thus functionalized in physical states or properties alone. Take the property of having a propositional attitude such as 'I believe that Mr. Wang is a nice guy' for example. Even though this property could be functionalizable, its functional specification has to involve other mental properties, which again need to be further analyzed. A single mental property could involve a whole web of mental properties, each of which needs to have its causal

[11]The exceptions would be people who argue for a form of dualism, such as Chalmers, and perhaps Jackson as well.

specification if it is to be functionalized. But not all of the causal specification can involve only sensory inputs and behavior outputs, or use only physical vocabulary. The functional/causal analysis may run in circles, and it is very likely that we may ultimately not be able to use only physical vocabulary to specify the function of a single mental property. Even if all mental properties are indeed physically realizable, it does not mean that we can thus specify them in purely physical terms. In other words, Kim has not established the move in his argument from the functionalizability of mental properties to the theoretical reducibility of mental properties; nor has he shown the reducibility of psychology to neuroscience. I thus conclude that Kim's model can only guarantee us ontological reduction; not intertheoretic reduction.

Kim thinks that the advantage his MTP has over Davidson's token physicalism is that it further explains how mental *types* are related to physical *types*. If one rejects the reduction thesis, then, so argues Kim, one has to come up with an explanation that satisfies the following conditions:

(i) it does not violate the physical closure principle;

(ii) it dissipates the causal competition between physical properties and mental properties;

(iii) it explains psychophysical type-type relations.

What Davidson's *anomalous monism* fails to provide, according to Kim, is an explanation that satisfies the third condition. In the remainder of this paper, I will try to delineate a relationship between mental and physical properties such that it does not deny that the physical domain is causally closed, and yet it also resolves the causal competition between mental and physical properties. As a nonreductionist, however, I will also want to show that mental properties are real, nonreducible properties and that they are also *causally relevant.*

3. A NONREDUCTIONIST'S SOLUTION: HOLISTIC IDENTITY

To resolve the tension between two competing causal explanations, using Kim's words, "we need to know how the two explanations are related, how the two causal stories about a single phenomenon mesh with each other. Are the two stories at bottom one story couched in different languages? Do the two stories supplement one another, each being only partial? And so on." (Kim (1998b), p. 66) In my analysis, the two causal stories would each be a sufficient and complete causal story, but there is no causal competition because they would turn out to be one story couched in different languages. In a nutshell, my solution is that of *identity*. But I shall argue that this kind of identity does not lead one to intertheoretic reductionism.

The best way to characterize the relationship between mental properties and physical properties is to use some version of *supervenience*. Kim argues that weak supervenience is too weak to be a determination relationship, and strong supervenience leads to reduction. We (the nonreductionists) seem to be in a bind here: either we give up the claim that the physical determine the mental, or we embrace the claim that the mental is reducible to the physical. Here I shall sketch an approach that does not deny strong supervenience and yet also does not accept the possibility of psychophysical reduction. I think the key is to specify which mental properties and which physical properties are involved in our formulation of *psychophysical supervenience*.

With Kim's *psychophysical supervenience*, we have this relationship:

[PS] Every internal psychological state of an organism is supervenient on its *synchronous* internal physical state. (Kim (1982), p. 183)[12]

[12]In a later paper, "'Strong" and "Global" Supervenience Revisited', Kim seems to extend this 'internal psychophysical supervenience' to a wide psychophysical supervenience. He says, "The lesson we learn from the Putnam and Burge-style cases that is relevant to the topic of supervenience is that

At the same time, with Kim's *strong supervenience*, we have the following relationship:

[SS] A *strongly supervenes* on B with respect to domain D just in case for any x and y in D, and any worlds w and w^*, if x in w is B-indiscernible from y in w^*, then x in w is A-indiscernible from y in w^*. (Kim (1988), p. 111)

What **[PS]** and **[SS]** state is that if two organisms x and y are "indiscernible" in their internal (*and* external or relational, under some interpretation) physical states, then they share the same psychological states. Now let us take a close look at the nature of these states.

The mental states Kim considers are all *post-instantiation* mental states; that is, they are the actual thoughts, beliefs, desires, etc., that some individual or another has had. For such an individual, the internal physical states that process these thoughts, beliefs are what Kim calls the "synchronous" internal (and external/relational) physical state. This supervenience relation is based on Davidson's token-identity theory. Davidson's token-identity theory states that singular mental events are identical with singular physical events, but there is no matching between mental event types and physical event types. Davidson proposes that we define 'event' as an individual happening that is spatially and temporally confined. Events are thus unrepeatable, dated individuals. Davidson thinks that what we need for asserting event-identity is a way to fill in the blank in "If x and y are events, then $x = y$ if and only if _____". He gives the following condition: "Many events are changes in a substance. If an event a is a change in some

contrary to what used to be taken for granted, many intentional states turn out not to be 'intrinsic' or 'internal' to the subjects to whom they are attributed.... It is no surprise then that we must seek a wider physical supervenience base for them, including relations and relational properties, if they are thought to be supervenient on the physical." (Kim (1993a), pp. 88-89.)

substance, then $a = b$ only if b is also a change in the same substance." (Davidson (1970a), p. 173) Thus, a mental event m can be identical to a neurophysiological event n only if both m and n occur in the same individual person. Having the spatial location of both events confined to the same person constitute a *necessary* condition, but not a *sufficient* condition, of the mental-physical event-identity. The second necessary condition needed is a temporal specification: "if events are identical, they constitute identical stretches of time." (Davidson (1970a), p. 177) But while spatial and temporal demarcations are useful for the individuation of events, they are insufficient in establishing the identity claim. What is needed to have an identity relation, according to Davidson, is an identity of the causal connection. Davidson's final thesis of token-identity can be summarized as this: *A mental event and a physical event are identical if and only if they happen to the same person at the same time, and they have exactly the same causes and effects*

However, under Davidson's token-identity thesis, there is still a problem for the individuation of neurophysiological events: even if we could grant that the spatiality of a neurophysiological event is easily defined as the individual's brain (or the smallest part of the brain where the change takes place), how do we define events in the brain *temporally*? Presumably, a neurophysiological event can be specified as whatever happens in the brain within an hour, a minute, a second, or even a millionth of a second. Which is *the event* we want to pick out as a neurophysiological token? It would seem that when we try to locate the neurophysiological token, what we are saying is that the neurophysiological correlate of one's having the belief that P at time t_1 is a relevant token, or that the neurophysiological correlate of one's having a pain experience at time t_2 is another relevant token. But when neurological events are specified this way, there are two problems. One is that the demarcation is still uncertain. For example, if the belief in question is a dispositional belief state not being presently thought about, what is its relevant neurophysiological correlate? Secondly, the

individuation of neurophysiological events is already based on a token-identity, and thus the individuation itself loses its explanatory status.

To avoid this trouble with Davidson's token-identity theory, I suggest that we abandon talk about "physical tokens" and use a *holistic* picture instead. Holism, to begin with, is the principle that a part is understandable only in its relation to the whole. In other words, no token can be locally, *atomically* defined. Therefore, if we want to understand each single token, we have to consider it not as a simple element, but as *a part* of the whole system. Two tokens of different wholes cannot be compared out of their respective systems since the tokens themselves are individuated only by the roles they play in each system. We can state this thesis as follows:[13]

[Holism] The individuation of x is determined by x's contribution to the whole system.

Under this thesis, a neurophysiological event is individuated by how it occurs in the whole brain of a particular individual and what effects it brings to that brain. A mental state such as a belief that P, on the other hand, is individuated by what else the person believes (in particular, her metalinguistic beliefs concerning the meaning of words involved in the specification of P). Mental holism has been widely discussed and has gained some endorsement. What I think has been neglected is the acknowledgment of "neural holism." For two individuals, x and y, to be indiscernible in their physical states, the physical state of the brain must be viewed not merely as one particular activity in the brain, but as a cross-section of the brain. For example, even when we suppose that there is an identifiable neural activity that

[13]To give a precise definition of 'holism' can be the most difficult task. There are many attempts to date and they may not all express the same idea. In this paper I shall be using the term 'holism' in a loose way as defined here.

corresponds to an individual's desire for coffee, the same neural activity occurring in person A may lead A to drink coffee while in person B it may not have the same effect. It is not just the neuron firing considered in isolation, but the neuron firing in that particular neural network, that plays a *deterministic* causal role in the person's action. We must therefore consider the whole neurological state of the brain, or we may call it the "neural map", in our formulation of the determinate causal history.[14]

A *holistic* specification of events is similar to, but not restricted to, a "causal-role" specification. The difference can be presented in these two respects: Firstly, it is not restricted to analyzing the immediate cause and effect of this particular event, but is looking at the whole history of the brain. Secondly, it allows us to consider the neuronal membrane *at rest*, or a dispositional mental state not presently affecting the individual's behavior.

Based on mental holism and neural holism, I propose that the mind/brain relationship be analyzed in terms of what I call *holistic identity*:

[14]This holistic treatment of the brain is supported by some empirical evidence provided by neurosciences. For example, studies show that the left hemisphere of the brain is dominant for speech in 96% of right-handed people and in 70% of left-handed people. If we only look at a particular temporally specified neurological event, without considering its contribution to the function of the whole brain, we might miss a crucial neurological difference between a left-handed and a right-handed person. Neuroscientists acknowledge that it is unlikely that we will ever be able to understand how each of the hundred billion neurons in our brain contributes to the function of the brain, but what they hope to accomplish is to explain how each category of neurons contributes to the whole system. Granting that one person's brain could be significantly different from another person's brain, we need to first secure the holistic identity between each individual's mind and brain before we move on to generalizations on either level.

[Holistic Identity] Let A be a psychological property *holistically specified*, let B be a neurophysiological property *holistically specified*; let x be any object, A is identical to B if and only if if x has A, it necessarily has B, and *vice versa*.

Identity, as construed in type-identity theories or token-identity theories, is not a notion about object identity, but about property identity. As Joseph Levine puts it, questions about property identity are "questions about how or why distinct properties are coinstantiated." (Levine (1998), p. 465) Type-identity theories specify the coinstantiated properties as psychological proper types and physical proper types. Token-identity theories specify the coinstantiated properties as properties of a spatially/temporally confined token/event. When properties are specified *holistically*, they are properties of tokens *considered in relation to a whole system to which the token belongs* By Fodor's definition, "Holistic properties are properties such that, if anything has them, then *lots* of other things must have them too" (Fodor & Lepore (1992), p. 2). Under this definition, the mind-brain identity relation holds only between one's mental state, considered in relation to one's whole mental life, and one's neurophysiological state, considered in relation to one's whole brain. Furthermore, when we consider neurophysiological events as tokens of a brain, the brain itself is not to be taken as a static whole. Since the brain is an organic whole, we cannot just demarcate its events *temporally*.[15] Each synchronic picture (each time-slice) of the brain can only be viewed as a frame in the whole dynamic system. In this kind of holistic picture, a part is related to the whole not as a unit in a well-constructed, interrelated "web", but as a unit of a stage in an ongoing process. A mental state such as a belief that P, on the other hand, is

[15]Putnam seems to be pushing for a similar picture back in 1973, when he wrote: 'Hull's model for rote learning suggests that some brain processes are best conceptualized in terms or continuous rather than discrete variables." (Putnam (1973), p. 139)

individuated by what else the person believes (in particular, her metalinguistic beliefs concerning the meaning of words involved in the specification of *P*), and what else the person desires, etc. These psychological characteristics of the individual also reflects the past history and personality traits of the individual, just as the individual's present brain state reflects what stimuli and inputs he or she acquired in the past. It is therefore practically impossible to find psychological twins that are not neurophysiological twins, and vice versa.[16]

Once the identity is established, we can resolve the exclusion problem. Following Davidson, I argue that mental events and physical events do not compete for causal power, since they are identical events. And under *holistic identity*, mental properties and physical properties also do not compete for explanatory relevance, since they are coinstantiated properties. But I want to emphasize that it is only the whole mental life of a person taken at a particular time-slice that has the determining causal power over what the person would think at the next time-slice. In the same way, it is only the whole neural map of a brain at the particular time-slice that has the determining causal power over the next neural map of the brain at the next time-slice. We can formulate the two causal explanations in this way:

Supposing that neurophysiological state N_1 leads to N_2, we have a *strict* neurophysiological causal law:

[NL] 'If x is in N_1, then x will be in N_2' or 'N_1 causes N_2'.

Now suppose N_1 is the neurophysiological base of M_1, and N_2 is the neurophysiological base of M_2. I argue that the following causal law also holds:

[16]In other words, I think that once psychological properties are specified holistically, they are not multiply realizable.

[ML] 'If x is in M_1, then x will be in M_2' or 'M_1 causes M_2'.

Both of these causal explanations give the complete causal stories and they are simply the same causal history given different descriptions. My picture of the mental/physical kind relationship is the following ('→' stands for causation, while '||' stands for identity):

M_1	→	M_2
||		||
N_1	→	N_2

Using the above identity relation, we can also formulate a *strong* supervenience relationship as follows:

[SS] 'If x is in N_1 and x is in M_1, when M_1 is a mental kind, then necessarily for all y's, if y is in N_1, then y is in M_1.'

In other words, the mental is determined by the physical and is thus not an "ontological dangler."

Given this holistic identity relationship, the whole mental life of an individual is *locally reducible to* the whole brain *extrinsically, relationally* considered.[17] By local reduction, I do not mean species-specific reduction, as some (including Kim himself) take it to mean. What I mean is a reduction restricted locally to one particular person. Kim thinks that we can have species-specific psychophysical identity and this identity serves as the foundation for species-specific psychophysical reductionism. But local reduction between an individual's mental life and her brain states does not warrant species-wide reduction between general

[17]I would call this kind of reduction '*wide* reduction' and this kind of mental/physical supervenience '*wide* supervenience.'

mental and physical properties. I argue that such a species-wide identity is not possible; furthermore, I argue that this local reduction is actually a disproof of intertheoretic reductionism.

To begin with, neither the mental kind M_1 nor the neurophysiological kind N_1 involved in this supervenience can be the proper kinds of psychology and neurophysiology *per se*. Kim calls the description of all the neurophysiological properties of this cross-section of the brain "the N-maximal properties." The neurophysiological kind N_1 is thus the set of all the neurophysiological properties of the brain. As we know, the neural map in our brains differs from one person to the next; it also changes constantly within the same person. There are about ten billion neurons in the cerebral cortex alone, and among these neurons, there are about one million billion neural connections in the cortical sheet.

> If we consider how connections might be variously combined, the number would be hyperastronomical – on the order of ten followed by millions of zeros…. And even more remarkable is the way in which brain cells are arranged in functioning patterns. When this exquisite arrangement of cells is taken together with the number of cells in an object the size of your brain, and when one considers the chemical reactions going on inside, one is talking about the most complicated material object in the known universe. (Edelman (1992), p. 17)

Therefore, no two neural states will be exactly alike in reality, even though we could find "molecule-for-molecule" identical brains in theory. This being the reality of the brain, there would be infinitely many neurophysiological kinds from N_1 to N_n. It is clear that neuroscience cannot use N_1 to N_n as its proper kinds. My speculation is that neurophysiology can only give a regional measurement or a coarse-grained description of each region's causal responsibility. We can call them *macro*-neurophysiological kinds. In contrast, we can call N_1 to N_n *micro*-neurophysiological kinds. With micro-neurophysiological properties, there is even a problem with

measurement: No neuroscientist can ever claim to give a precise measurement of all the neural activities of every neuron in the brain. If *macro*-neurophysiological kinds cannot be identified with *micro*-neurophysiological kinds, then local physical realization does not warrant a species-wide unified physical realization. Ultimately, what neurophysiology taxonomizes as kinds are not what we are describing as N_1 to N_n – the fine-grained causally determining kinds.

In the same way, the mental kind that supervenes on this micro-neurophysiological kind is not a kind that we have in the taxonomy of psychology *per se*, but a much more finer-grained taxonomy of what I will call '*individual-specific* psychology,' or '*IPsy*.' *IPsy* has predicates of the form 'x has the mental state that x believes that p and x desires that q and that.... (which exhausts all of x's mental inclinations that are relevant to the decision).' For example, both x and y love crabs and they are both staying at a hotel which is fourteen blocks away from a famous Crab House. A general intentional psychological explanation would be giving the following generalization: 'If x loves crabs and x believes that at the Crab House x can get crabs, then x would go to the Crab House.' It is clear that such a generalization cannot be strict: there are too many exceptions. But now if we specify all the mental inclinations that come into play: such as, both x and y hate walking fourteen blocks in the evening; both x and y think about taking a taxi to go to the Crab House; however, x does not mind taking a taxi while y has a deep fear about riding in a taxi by herself, etc., then we would put x's and y's mental states prior to their decisions into different *IPsy*-kinds. We can also explain why x ends up going to the Crab House while y ends up not going.

Each *IPsy*-kind includes the complete description of the *total* instantaneous states of the individual – states which are described as the total specification of the present condition of that individual. Such a mental state determines, together with learning and memory, what the next state will be. This taxonomy gives a *holistic* consideration of

one's mental life, taking the present mental life as the end-state of previous learning experiences as well as memory. This *IPsy* taxonomy is of course not very useful to us, since we cannot generalize from one person to the next. But theoretically, any other individual who shares the *IPsy*-state with *x* would just be taking the same action as *x* would.

Now I argue that the M_1 and M_2 that correspond to the N_1 and N_2 in the neurological causal law [NL]: N_1 causes N_2, are *IPsy* -kinds. And given this understanding, we can see that both [NL] and [ML] can be strict laws, and yet neither has any generalizability *in practice*. Any actual neurophysiological law *that is to be strict* can have only one incident, involving only one neural map of a particular brain at a time. We may call it a *token-specific* law. Similarly, psychological laws can also be strict – as long as they are specified in terms of *IPsy*-kinds and are *token-specific*. In other words, if the neurological state (specified by a set of N-maximal properties) prior to an individual's taking an action is *causally sufficient* in bringing about this action, then the mental state (specified by a set of *IPsy*-maximal properties) corresponding to this neurological state is equally *causally sufficient* in bringing about the action. If there is a strict physical law linking the neurological state to the action, so is there a strict mental law linking this individual-specific psychological state to the action. On the token-specific, post-instantiation level, the two sets of "maximal properties" are identical and are mutually reducible. With the issue of causal exclusion, I thus conclude that there is no causal competition between *IPsy*-maximal properties and N-maximal properties. If it is true that (under the supervenience thesis), any *y* that is in N_1 must be in $IPsy_1$, and any *y* that is in N_2 must be in $IPsy_2$, then when there is a causal strictness between N_1 and N_2, there must be a causal strictness between $IPsy_1$ and $IPsy_2$.

But acknowledging this identification would not lead us to embrace the reduction of general psychology to general neurophysiology. We shall now establish the relationship between general psychology and general neurophysiology on the basis of the

relationship between *individual-specific* psychology and *micro* neurophysiology. To begin with, the connection between *IPsy* and general psychology is more easily established: Psychology could be viewed as a generalization of these individual-specific psychologies. Psychological taxonomy is not based on the grouping of the whole mental life of every individual, but on the grouping of what different mental lives share in common. Now let us consider psychology *per se*. Psychology (*Psy*) and *IPsy* are two sciences at different levels of structure and organization. *Psy* is a generalization based on *IPsy*, or we should say, enough cases of *IPsy*. So, for each *Psy*-type, we would have numerous *IPsy*-types. This *Psy* -type is the result of generalization of numerous *IPsy*-types, and guided by our purpose or our interest in study, we select some common aspects of the *IPsy*-types as relevant in our individuation of *Psy*-types:

$$Psy_i = \{IPsy_1 \vee IPsy_2 \vee IPsy_3 \ldots\ldots IPsy_n\}$$

After the generalization process, the causally determining power is lost. When the mental state at issue is formulated as a *Psy*-type, a psychological law using it as the causal antecedent must add a "ceteris paribus" clause:

[PL] If Psy_i, then action$_j$ (or another mental state characterizable as Psy_j), *ceteris paribus*.

Some of the psychological factors included in the *ceteris paribus* clause may actually be a determinate "countervailing" factor in some people, even though in most people they may be insignificant. This explains why general psychological laws are not *strict*.

The connection between each *Psy*-type and *IPsy*-type is explained as above. But on the other hand, the connection between the each *micro* neurophysiological type and general neurophysiological type is not so clear.

We can show how different *IPsy*-types $\{IPsy_1 \vee IPsy_2 \vee IPsy_3 \ldots\ldots IPsy_n\}$ could be generalized into a psychological type Psy_i, because we can use the same mental predicates in both taxonomies. But we have no evidence that there is any generalizability between N_1 and N_2, ... and N_n. How neurophysiology generates types is a more complicated issue because variations in single neuron firings, in neural patterns, or in neurotransmitter systems, etc. could be infinitely great. I suspect that neuroscience can only describe regional responsibility *coarsely* defined. That is to say, it may be in principle possible for neuroscientists to tell us what part in the brain is responsible for language, for sensation, etc., but it may never be possible for them to give us a complete description of the N-maximal properties that underlie a person's every mental activity however minutely individuated. From how vastly different these individual neurophysiological bases underlying the same *IPsy*-kind can be, we simply do not see the possibility of their being generalized into *one* neurophysiological kind.

We can now see why the type-type identification between general mental properties and physical properties is lost. Each psychological type has an indefinite, or even infinite, number of actual and potential neurophysiological types as its physical bases. This is not just the problem of *multiple realizability*. The problem is rather that in our *bottom-up* generalization from individual-specific psychology to general psychology, many mental details are left out for the sake of grouping. But it is the detailed, fine-grained, micro properties that have one-to-one psychophysical correspondence. Furthermore, once properties are specified *holistically*, they are confined in the system to which they belong. On the level of mental properties, a particular mental property is specified in relation to other mental properties. On the level of neurophysiological properties, a particular neurophysiological property is specified in relation to other neurophysiological properties. Each level has its own *web* of interconnections and there is little evidence that we can match the

psychological map and the neural map exactly. Therefore, the local reduction that I favor is at most a *property-reduction* between mental properties (specified individual-psychologically) and micro-neurophysiological properties (specified as a complete description of the whole neural network of one individual). It is not a *theory-reduction* between psychology and neurophysiology. That there are neurophysiological state does not imply that there are *neurophysiologically taxomizable* states, just like the fact that there are psychological states does not mean that there are *psychologically taxomizable* states. The mental properties that are reducible to neurophysiological properties do not constitute a (general) psychological kind. The neurophysiological bases of this mental kind also do not constitute a (general) neurophysiological kind. Intertheoretic reduction between psychology and neurophysiology is not feasible, but this does not entail that the mental realm does not have the physical realm as its base. *What I deny is merely the reductionism between two special sciences, not the reduction between two metaphysical realms.*

If general psychology is not reducible to general neurophysiology, then the explanatory competition between the *two theories* still exists. So we are back to the explanatory exclusion problem that Kim advocates. My solution is twofold. First of all, Kim's explanatory exclusion problem is based on his *Causal Explanatory Principle* [CEP], but the kind of causal explanation that Kim demands through this principle can only be offered either an *individual-specific* psychological explanation, or as a *micro*-neurophysiological explanation. These explanations are what I formulated as [ML] and [NL] respectively. Therefore, on the local, individual-specific level, the two explanations do not compete. Secondly, on the level of general psychological explanation and general neurophysiological explanation (applied species-wide to human beings), Kim's [CEP] does not apply since neither theory can offer *strict* causal laws. We therefore do not have to choose one theory over the other since one general explanation does not necessarily exclude the other.

Furthermore, I suspect that neuroscience can never offer any useful explanation for human behavior without appealing to some intentional psychological explanation. To have a neurophysiological measurement of the causal history of a behavior, we need to be able to interpret the results of our measurement. And the interpretation of a neurophysiological measurement has to rely on psychological interpretation of the behavior. In order for neuroscientists to find regional responsibility, they would also have to measure (with whatever new techniques they will produce in the future) the neural *action* states and *rest* states corresponding to a known mental phenomenon. In other words, neuroscientists do not have a way of defining the relevant neural correlate *if they do not already assume a psycho-physical identity.* The explanatory power of neuroscience will ultimately rest on the ontological model of psycho-physical identity. Mental properties do not become causally/explanatorily irrelevant either when they are specified relative to each individual, or when they are specified as general psychological properties. If so, then there is another way out of the explanatory exclusion: Granting that psychological explanation and neurophysiological explanation exclude each other, we see that what should be excluded is the neurophysiological explanation, not the psychological explanation.

4. CONCLUSION

In this paper I have tried to argue against Kim's view that the only solution to the exclusion problem is that of reductionism. I have demonstrated a way to link the mental with the physical such that the mental is not causally inert, and at the same time psychology is not made to be reducible to neurophysiology. I want to emphasize again that my position *is* for the ontological reduction from the mental to the physical. But to establish intertheoretic reduction between psychology and neuroscience, we need to have the incentive that the latter theory will be able to give a simpler and yet fuller explanation than the former

does. Putnam says, "It seems to me that whatever the pragmatic constraints on explanation may or may not be, one constraint is surely this: That the relevant features of a situation should be brought out by an explanation and not buried in a mass of irrelevant information." (Putnam (1973), p. 132) With the massive details on the micro-neuro-physical level, neuroscience alone will not be able to give us the information that we are interested in seeking on the level of intentional psychology.

Finally, I want to point out that in this paper I am not just offering a nonreductionist's alternative to Kim's reductionistic solution, but a rejection of his reductionism. Kim's argument for reduction is based on the existence of biconditional correlation between a mental property M and a disjunctive set of M's physical bases, N_1 or N_2 ... or N_n, such that $M \leftrightarrow N_1 \vee N_2 \vee ... \vee N_n$. A major problem with this biconditional is that, the disjunctive set of physical bases $\{N_1 \vee N_2 \vee ... \vee N_n\}$ is too large to be made use of in offering causal explanations. The problem is not just the multiple realizability among different species or different biological/mechanical structures. As Kim sees it, we can be just talking about species-specific reduction. The problem is rather that even within the same species such as human beings, and even within the same individual, the variations of physical instantiations of a mental whole are infinitely many. As Terence Horgan puts it:

> For all we now know ... the intentional mental states we attribute to one another might turn out to be radically multiply realizable at the neurobiological level of description, even in humans; ... indeed, even in an individual human *given the structure of his central nervous system at a single moment of his life.* A single person S at a single moment t might be capable of undergoing (at t) any of a variety of different physico-chemical states which are quite disparate from one another at the physico-chemical level of description ... each of which would realize (in S) the same intentional mental property. (Horgan (1993), p. 308.)

Since the kind of deterministic mental state is a conjunctive set of all the individual's mental attributes relevant to the causation of her action, the physical base of this mental kind must itself be a conjunctive set of all the micro-physiological properties of the individual's brain. Given the variations of each individual's brains in their minutest details, the disjunctive set of all the possible physical bases is inexhaustible. Furthermore, given that the physical bases are bases of "instantiated" mental states, they could even compose an *open* infinite set since the mental states at issue could still be instantiated at any time to any individual in the future. The same criticism Putnam has toward the Turing machine states theory can be used against the neurophysiological states theory: "Besides the fact that such a description would be literally infinite, the theory is now without content..." (Putnam (1973), p. 137)

REFERENCES

ANTONY, L. (1989). "Anomalous Monism and the Problem of Explanatory Force", *The Philosophical Review*, Vol. XCVIII, No. 2, pp. 153-187.

————. (1995). "I'm a Mother, I Worry", *Philosophical Issues*, 6, *Content* (ed.) Enrique Villanueva, (Ridgeview Publishing Company, Atascadero CA) pp. 160-166.

ANTONY, L. & LEVINE, J. (1997). "Reduction with Autonomy", *Philosophical Perspectives* 11, *Mind, Causation and World*, pp. 83-105.

BLOCK, N. (1990). "Can the Mind Change the World?" in *Meaning and Method: Essays in Honor of Hilary Putnam*, George Boolos (ed.) (Cambridge, Cambridge University Press) pp. 137-170.

————. (1997), "Anti-Reductionism Slaps Back", (manuscript).

BURGE, T. (1993). "Mind-Body Causation and Explanatory Practice", in *Mental Causation*, John Heil & Alfred Mele (eds.), (Oxford, Clarendon Press), pp. 97 - 120.

CAMPOS, C. (1995). "Kim on the Exclusion Problem", *Philosophical Issues*, 6, *Content*, Enrique Villanueva (ed.), (Ridgeview Publishing Company, Atascadero CA) pp. 167-170.

CARTWRIGHT, N. (1983). *How the Laws of Physics Lie*, (New York, Oxford University Press).

DAVIDSON, D. (1967). "Causal Relations", reprinted in Davidson (1980), pp. 9-162.

————. (1970a). "Events as Particulars", reprinted in Davidson (1980), pp. 181-187.

————. (1970b). "Mental Events", reprinted in Davidson (1980), pp. 207-227.

————. (1975). "Causal Relations", in *Causation and Conditionals*, (ed.) Ernest Sosa, (Oxford University Press, London), pp. 82-94.

————. (1976). "Hempel on Explaining Action", reprinted in Davidson (1980), pp. 261-276.

————. (1980). *Essays on Actions and Events*, (New York, Oxford University Press).

————. (1993). "Thinking Causes", in Heil & Mele, (1993), pp. 3- 17.

DAVID, M. (1997). "Kim's Functionalism", *Philosophical Perspectives* 11: *Mind, Causation and World*, pp. 133-8.

DRETSKE, F. (1989). "Reasons and Causes", in *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, pp. 1-15.

————. (1995). *Explaining Behavior: Reasons in a World of Causes*, (Cambridge, Mass., MIT Press).

EDELMAN, G.M. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind*, (Basic Books, A Division of Harper/Collins Publishers, Inc.).

FODOR, J. (1974). "Special Sciences", Reprinted in Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science*, (Cambridge, Mass., MIT Press), (1981), pp. 127-5.

————. (1997). "Special Sciences: Still Autonomous After All These Years", *Philosophical Perspectives* 11: *Mind, Causation and World*, pp. 9-163.

FODOR, J. & LEPORE, E. (1992). *Holism: A Shopper's Guide* (Blackwell, Oxford).

HEIL, J. & MELE, A. (eds.) (1993). *Mental Causation*, (Oxford, Clarendon Press).

HORGAN, T. (1989). "Mental Quausation", *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, pp. 47-76.

————. (1993). "Nonreductive Materialism and the Explanatory Autonomy of Psychology", in *Naturalism: a Critical Appraisal*, (eds.) Steven Wagner & Richard Warner, (Notre Dame, University of Notre Dame Press).

————. (1997). "Kim on Mental Causation and Causal Exclusion", *Philosophical Perspectives* 11: *Mind, Causation and World*, pp. 165-184.

JACKSON, F. & PETIT, P. (1990a). "Program Explanation: A General Perspective", *Analysis*, Vol. 50, pp. 107-117.

————. (1990b). "Causation in the Philosophy of Mind", *Philosophy and Phenomenological Research*, vol. L, Supplement, pp. 195-2.

KIM, J. (1973a). "Causation, Nomic Subsumption, and the Concept of Event", *Journal of Philosophy*, vol. LXX, No. 8, pp. 217-236.

————. (1973b). "Causes and Counterfactuals", *Journal of Philosophy*, vol. LXX, pp. 570-572.

————. (1979). "Causality, Identity, and Supervenience in the Mind-Body Problem", *Midwest Studies in Philosophy*, IV, pp. 31-49.

————. (1982). "Psychophysical Supervenience", in Kim (1993a), pp. 175-193.

————. (1983). "Supervenience and Supervenient Causation", *Southern Journal of Philosophy*, supp. Vol. 22, pp. 45-55.

————. (1984). "Epiphenomenal and Supervenient Causation", reprinted in Kim (1993a), pp. 92-108.

————. (1987a). "Explanatory Realism, Causal Realism, and Explanatory Exclusion", in *Midwest Studies in Philosophy*, XII, P. French, *et al.* (eds.), (Notre Dame, University of Notre Dame Press), pp. 225-239.

————. (1987b). "'Strong' and 'Global' Supervenience Revisited", reprinted in Kim (1993a), pp. 79-91.

————. (1988). "Supervenience for Multiple Domains", reprinted in Kim (1993a), pp. 109-130.

————. (1989a). "Mechanism, Purpose, and Explanatory Exclusion", reprinted in Kim, (1993a), pp. 237-264.

————. (1989b). "The Myth of Nonreductive Materialism", reprinted in Kim, (1993a), pp. 265-284.

————. (1990). "Explanatory Exclusion and the Problem of Mental Causation", reprinted in *Philosophy of Psychology: Debates on Psychological Explanation*, Cynthia MacDonald (ed.), (Cambridge MA, Blackwell, 1995), pp. 121-1.

————. (1991). "Dretske on How Reasons Explain Behavior", reprinted in Kim (1993a), pp. 285-308.

————. (1992). " 'Downward Causation' in Emergentism and Non-reductive Physicalism", In *Emergence or Reductionism?*, A. Beckermann, H. Flohr, and J. Kim (eds.), (Berlin, De Gruyter), pp. 119-138.

———. (1993a). *Supervenience and Mind: Selected Philosophical Essays* (Cambridge, Cambridge University Press).

———. (1993b). "The Nonreductivist's Troubles with Mental Causation", reprinted in Kim, (1993a), pp. 336-357.

———. (1993c). "Postscripts on Mental Causation", reprinted in Kim, (1993a), pp. 358-367.

———. (1993d). "Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?", in Heil & Mele (eds.), (1993), pp. 19-26.

———. (1994). "Explanatory Knowledge and Metaphysical Dependence", *Philosophical Issues*, Vol. 5, *Truth and Rationality* E. Villanueva (ed.) (Ridgeview Publishing Company, Atascadero CA).

———. (1995). "Mental Causation: What? Me Worry?" in *Content: Philosophical Issues*, 6, Enrique Villanueva (ed.), (Atascadero, California, Ridgeview Publishing Company), pp. 123-151.

———. (1996). *Philosophy of Mind*, (Boulder, Westview Press).

———. (1998a). "The Mind-Body Problem After Fifty Years", in *Current Issues in Philosophy of Mind*, (ed.) Anthony O'Hear, (Cambridge, England, Cambridge University Press), pp. 3-21.

———. (1998b). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, (Cambridge, Mass., MIT Press).

———. (1999) "Mental Causation and Consciousness: The Two Mind-Body Problems for the Physicalists", (manuscript).

LEVINE, J. (1998). "Conceivability and the Metaphysics of Mind", *Nous* 32:4, pp. 449-480.

LEWIS, D. (1986). "Causation", in David Lewis, *Philosophical Papers*, vol. II, (Oxford, Oxford University Press, 1987), pp. 159-173.

MELNYK, A. (1994). "Being A Physicalist: How and (More Importantly) Why", *Philosophical Studies* 74, pp. 221-241.

———. (1995). "Two Cheers for Reductionism: Or, the Dim Prospect for Non-Reductive Materialism", *Philosophy of Science* 62, pp. 370-388.

PUTNAM, H. (1973). "Reductionism and the Nature of Psychology", *Cognition* 2(1), pp. 131-6.

SABATÉS, M.H. (1996a). "Kim on the Metaphysics of Explanation", *Manuscrito*, Vol. XIX, No. 2, pp. 93-110.

———. (1996b). "Should A Cognitive Psychologist Worry About the Causal Inefficacy of the Mental?" (Manuscript).

SEARLE, J. (1992). "Reductionism and the Irreducibility of Consciousness", from *The Rediscovery of the Mind*, Chap. 5, reprinted in *The Nature of Consciousness*, (eds.) Ned Block, Owen Flanagan, and Güven Güzeldere, (Cambridge, Mass., MIT Press), pp. 451-459.

SOSA, E. (1984). "Mind-Body Interaction and Supervenient Causation", in P.A. French *et al.* (eds.), *Midwest Studies in Philosophy*, IX. (Minneapolis, University of Minnesota Press), pp. 271-281.

TELLER, P. (1983). "Comments on Kim's Paper", *Southern Journal of Philosophy*, Supp. Vol. 22, pp. 57-61.

TOMBERLIN, J. (1995). "Mental Causation: A Query for Kim", in *Philosophical Issues*, 6, *Content* Enrique Villanueva (ed.), (Ridgeview Publishing Company, Atascadero CA), pp. 152-159.

YABLO, S. (1992). "Mental Causation", *The Philosophical Review*, Vol. 101, No. 2, pp. 245-280.

———. (1997). "Wide Causation", *Philosophical Perspectives* 11: *Mind, Causation and World*, pp. 251-281.