

# Irreducibility and Subjectivity

Ron McClamrock  
University at Albany, SUNY

In *Philosophical Studies* 1992

...the problem of...how cognition...is possible at all...can never be answered on the basis of a prior knowledge of the transcendent [i.e. the external, spatio-temporal, empirical]...no matter whence the knowledge or the judgments are borrowed, not even if they are taken from the exact sciences.... It will not do to draw conclusions from existences of which one knows but which one cannot "see". "Seeing" does not lend itself to demonstration or deduction. [Husserl 1964a, pp. 2-3]

In the wave of materialistic philosophizing about the mind that has dominated the last 30 years, a persistent thorn in the side of the materialists has been the issue of subjectivity. For many, Thomas Nagel's "What is it like to be a bat?" has codified the intuition that no objective account of mind could adequately capture subjective phenomena. But Nagel's anti-objectivism has recently come under direct attack by several philosophers, notably including William Lycan [1988]. Although I'm sympathetic to Lycan's motivations and materialist preconceptions, I think that he's overstated the case against a view like Nagel's. In explaining the ways in which his criticism fails, I'll suggest that the most important point of a position like Nagel's is its *anti-reductionism* about the subjective rather than any kind of *anti-materialism*. I will then examine what kind of irreducible subjectivity might be possible even within a materialist framework.

## 1. Nagel on subjectivity

Most centrally, Nagel's article presents a conception of subjectivity which seems to many to capture something at the heart of the idea of an *irreducibly subjective character of experience*. As he puts it, "an organism has conscious mental states if and only if there is something it is like to *be* that organism --- something it is like *for* the organism." [Nagel 1979, p. 166] And it is this notion of subjectivity which he thinks stands squarely in the way of providing an objectifying (or "reductive") analysis of mind. He suggests that we have no way of seeing how such an analysis is even *possible*, since "any reductionist program has to be based on an analysis of what is to be reduced ... . Without some idea, therefore, of what the subjective character of experience is, we cannot know what is required of physicalist theory" [Nagel, p. 167]. Further, if an objective analysis of mind must account for the subjective character of experience, "we must admit that no presently available conception gives us a clue how this could be done" [Nagel, p. 175]. That is, we don't only lack a particular analysis of the mind-body relationship that can account for subjectivity; we furthermore lack an answer even to the question of how such a thing might be *possible* --- we lack even any kind of *schema* for an answer. But what would it be to have such a schema, if the traditional answers to the mind-body question --- e.g. behaviorism, functionalism, physicalism, etc. --- don't count as giving one? What is it to "leave out" subjectivity?

One anti-materialist (or anti-functionalist) line which Nagel does not centrally rely on is the appeal to the standard "zombie" intuition. Still, he doesn't hesitate to advocate it along the way. As he puts it, subjectivity is "not captured by any of the familiar, recently devised reductive analyses of the mental, for *all of them are logically compatible with its absence of subjective*

*character of experience*. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since they could be ascribed to robots of automata that behaved like people though they experienced nothing." [Nagel, pp. 166-7] This answer has surely been around forever [Consider, for example, Leibniz's version from *The Monadology*: "... perception and that which depends on it are *inexplicable by mechanical causes* ... supposing that there were a machine so constructed as to think, feel and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push one against another, but never anything by which to explain a perception" [Leibniz 1962, p. 206].] and it never really goes totally out of style. But it has seemed especially popular --- at least as a subject of debate --- over the last 20 years or so. John Searle's infamous "Chinese room" example [Searle 1980] and the various anti-functional homunculi-head examples [see Block 1978] are among the better-known of these. But the problems with this line are well-known. Most importantly, the intuition that the standard simulations of human structure *don't* have mentality, intentionality, subjectivity, or whatever it is that you want is not one that we all find so completely compelling. After all, if you weren't already pretty sure about the subjectivity lurking within human brains, you'd likely be pretty skeptical that this rumpled lump of hydrocarbons is what turns the cosmic trick.

Why is *subjectivity* in particular that is so hard to capture? As Nagel puts it: "The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view" [Nagel, p. 167]. This is the heart of his claim: In giving an account of subjectivity, one must capture the sense of a "point of view", and this is seen as impossible. Subjectivity is essentially tied to this idea of a point of view, but any objective analysis must leave that behind. There are facts (about "what it's like") which "embody" and "are accessible only from" [Nagel, p. 163] a particular point of view, and hence "cannot be captured by any objective reduction". [It's important that Nagel doesn't want his claim to be one about essentially *private* experience. We can perfectly well, on his view, imagine what it's like to be another human who is sufficiently like we are --- a friend, maybe even the vice-president.] So, for example, we can't imagine what it's like to be a bat, and to perceive by bat sonar, even though we might well think that they have some subjective character of experience. Or as Husserl put the same idea: "A man born deaf knows that there are sounds, that sounds produce harmonies and that a splendid art depends upon them. But he cannot understand *how* sounds do this, how musical compositions are possible. Such things he cannot *imagine*, i.e., he cannot "see" and in "seeing" grasp the "how" of such things ... ." [Husserl 1964a, pp. 3-4]

## 2. Lycan's response

In his reply to Nagel [Lycan, ch.7], Lycan allows that there is one natural reading of Nagel's point which is quite correct. Lycan readily acknowledges that "seeing someone's brain in a state of sensing-blazing-red is nothing at all like sensing blazing red oneself" [Lycan, p. 76]. Similarly with respect to the bat's sonar sensation *S*: "We do not know what it is like to have *S* (we do not have cognitive access to *S*) *in the way the bat does*."

But these facts, he says, are "welcomed by the materialist". When I observe the bat in some physical or functional state, I don't thereby *have* that state myself, and I don't have the same perspective with respect to it. But a materialist account of the mental should not claim otherwise;

as he puts it, "the felt incongruity is just what anyone, materialist or anti-materialist alike, should expect. Therefore the incongruity affords no objection whatever to materialism" [Lycan, p. 77]. The feeling of profound incongruity between *having* the state and *knowing from the outside about it* is a natural and expectable consequence of the fact that the two involve --- to adopt the functionalist/computationalist form --- representing the event in entirely distinct representational formats. As Paul Churchland [1988] puts essentially the same point: The critical difference "may reside not in *what* is respectively known by each (brain states by the former, nonphysical *qualia* by the latter), but rather in the different *type*, or *medium*, or *level* of representation each has of exactly the same thing: brain states ... ." [Churchland, p. 34].

This "functional" reply (as Lycan calls it) thus aims to *defuse* rather than *refute* the intuition of incongruity, and is quite plausible when directed against one common way of taking Nagel's claim. Suggesting that there are constraints and limits on the character of our representation of stimuli hardly constitutes giving an argument against (in particular, a roughly functionalist) materialism. Such limitations are *exactly* what one would expect on the functionalist or materialist account. As such, it's hardly plausible to see them as evidence against the views. There are ways of knowing or "grasping" you can't engage in without the right perspective, but this alone does not impugn materialism or functionalism.

But Lycan quite rightly sees that this response does not completely dispose of Nagel's claim. He acknowledges a second possible kind of essentially "non-objective" or "perspectival" fact -- facts that can only be *referred to* by making use of the appropriate subjective perspective. The challenge provided here is in my view more problematic; and Lycan's arguments against it less watertight.

According to Lycan, "we can suppose that [the fact's] alleged perspectivalness is located either in an individual constituent or in a property constituent." [Lycan, p. 79] For the sake of clarity, let me focus on the latter case. What would it be for a fact to have an intrinsically perspectival property as constituent? On Lycan's view, this amounts to claiming that "there is [a property] concept that can be grasped and/or reported only in a first-person, perspectival way, and not in the third-person, objective way." [Lycan, p. 79] But such a concept is a "function from worlds to sets of individuals" as he sees it; and he insists that "*any such function is objectively describable*, or so it would seem ... ." [Lycan, p. 79] The problem then is that "It seems Nagel will have to eschew this powerful and effective way of representing the constituents of propositions and facts if he is to maintain the existence of perspectival ones, and that we should be loath to do." [Lycan, p. 79]

Now on the whole, I'm significantly less loath to give up the "power and effectiveness" of possible worlds semantics than Lycan is. More 'situated' accounts of the semantics of propositions (like *situation semantics* [Barwise & Perry 1983]) aren't just to be ruled out. But even within this perspective, the argument here doesn't do the job. What *is* it to claim that there will be, for any allegedly perspectival property, an "objectively describable" or "not intrinsically perspectival" function from worlds to sets of individuals?

Lycan's gloss on the notion is slightly puzzling: "... there is nothing intrinsically perspectival about functions from worlds to individuals; any one could be described by anyone who had the right sort of mental apparatus or brain writing" [Lycan, p. 79]. But if this is what "objective" amounts to here, it's not clear how it answers Nagel at all. Nagel's claim was that there are facts such that knowledge of them (or perhaps even reference to their constituents) requires making

use of a particular sort of perspective --- such as that of being constituted in a certain way. If the sense in which the "function" is objectively describable is just that *anybody constituted in the right way* can characterize it, then no rebuttal has been given to the claim that no objective theory can characterize the facts about bat phenomenology so that they can be referred to by beings (like us) who don't share the bat's structure and perspective.

Part of the confusion here may come from Lycan's implicit dichotomy between reporting in a 'first-person, perspectival' way and in a 'third-person, objective' way. Nagel quite rightly allows that we might refer in the third person to intrinsically perspectival facts that we don't stand in exactly the right first-person relationship to --- as when I refer to your pains, or to Wade Boggs' perception of a curve ball. I'm surely doing so in some kind of mediate way --- by making critical use of someone *else's* perspective. But referring to the subjective state through the reports and authority of the subject hardly makes them non-subjective; I refer to them then in a third-person but still perspectival way. [Nagel of course does not think subjective properties are essentially private, just that constraints on one's mental life can make some *types* of them inaccessible [See Nagel, pp. 171-172].]

Furthermore, it's now an entirely ubiquitous suggestion that what we refer to is conditioned by the environmental context in which our thoughts and utterances occur. In particular, the case has been made *ad nauseam* for the natural-kind terms like 'water' and 'arthritis'; but the moral has been broadened by many --- especially by those working on the general topic of 'deictic' reference in linguistics. What fixes the reference of our words and thoughts isn't just "in the head", but in the structures of the physical world and social context around us. By making use of the context, there may be all sorts of ways to refer to features of the character of experience of beings other than ourselves. The constituents of "funny facts" needn't be entirely inaccessible in order for them to be undecidable in the language of any particular functional account of the mental.

### 3. The autonomy of the phenomenological

But the problems involved in resisting 'perspectival facts' go even deeper than this. Not only does Lycan's own argument fail, but there are also substantive reasons from within the philosophy of science that would seem to support at least some aspect of Nagel's claim.

Consider the standard functionalist move against a more reductive materialism --- a move which Lycan not only endorses, but characterizes as "the only positive doctrine in all of philosophy that I am prepared (if not licensed) to kill for" [Lycan, p. 37]. The standard token physicalist account of materialism (which goes hand-in-hand with functionalism) is motivated first and foremost by the sense that higher-level properties of systems --- functional properties, if you like --- are not *definable* in physical terms. The moral here is now commonplace: [See Davidson 1978; also McClamrock (forthcoming), chapter 1.] Token physicalism does not require property identity between the higher level and the lower-level --- in this case, the mental and the physical. This fundamental irreducibility makes the presence of higher-level sciences not just an accident of local epistemology, but a basic part of carving the world at its joints. As Fodor nicely puts it, the generalizations of the higher-level sciences "can often be made about events whose physical descriptions have nothing in common ... [and] it is often the case that *whether* the physical descriptions of the events subsumed by such generalizations have anything in common is, in an obvious sense, entirely irrelevant ... ." [Fodor 1975, p. 15] So, to use a standard example: To be a gene, or a gene of a particular type, is not the same thing as being a DNA molecule of a

particular type. They are not identical *properties*, although states of affairs in the world consisting in something being a gene of a certain type are also states of affairs consisting in something being a DNA molecule of a certain type. [See McClamrock 1991]

This suggests that if we were to count facts as distinct when they have distinct *property* constituents, then the fact that's known when you know that something is an instance of C-fiber stimulation and the fact that's known when you know something is an instance of pain are different facts. Lycan has claimed --- quite reasonably, as I've said --- that the "functional" account of "perspectival facts" doesn't make any headway against materialism or functionalism since the restrictions on knowledge generated are just what the materialist or functionalist should have expected, and is in no way in conflict with the view. But the present case is different. Here, the problem is not just that knowing the imagined functional account doesn't mean knowing what the experience is like from the standpoint of the experiencer; it's that knowing the functional account may not even facilitate *reference* to the experiential properties *as those kinds of phenomenological properties* --- just as physical descriptions don't pick out functional properties *as such*.

What is needed to refer to phenomenological *properties* as such in objective terms? Referring to the brain states that are token identical with the phenomenological events or states is not enough to accomplish picking out the appropriate phenomenological properties as such. After all, it's not adequate for picking out the properties of the system at the functional or computational level either --- that is the central point of the ubiquitous advocacies of token over type physicalism. Computational properties are essentially undefinable in the language of physics, or neurophysiology, for that matter. Reference to the property at the appropriate level has to organize and taxonomize it correctly. But it is simply an open possibility that no theory will do this. No physical theory of the brain will allow picking out computational properties *as such*; in the same way, it may well turn out that no computational account of the mind's working will allow defining the phenomenological properties of the brain as such.

## 4. Irreducibility revisited

The possibility for the autonomy of the phenomenological I've illustrated here, although not in the end anti-*materialist*, is strongly enough anti-*reductionist* to be of some concern to mainstream physicalists (like Churchland) and functionalists (like Lycan). On this possibility, the irreducibility of the subjective would be due, not to some general anti-materialism, but to the possible anomalousness of the phenomenological with respect to the computational. But then I don't think that anomalousness and autonomy between explanatory accounts should *ever* be seen as particularly surprising. Rather, it's the converse which should surprise us when it turns out to be the case. The relationship between higher-level and lower-level properties of complex systems is not just in *principle* multiple realizability; we may often find multiple realization of higher-level properties within the very same structure, and even multiple realization of the very same token higher-level entity.

This is particularly so when we enter domains in which something like a computational account of activity starts to seem appealing --- as it does here for Lycan. The implementation of higher-level primitives in computer programs provides a particularly clear example of the possibilities of multiple realizability even *within* a particular complex system. Consider a particular variable in a program; call it *VAR*. At different moments in the running of the program, the machine-level implementation of the primitive *VAR* may be quite different. It will, of course, hold different

values at different times. It also will reside at different real memory locations (as garbage collection may re-allocate variable memory), and so on. What makes all these implementations of *VAR* an important and interesting class of things is that they play a fundamental and reliable role in the processes of the system at the higher level of organization. They may not share any terribly interesting properties at the machine level that they don't also share with, say, implementations of other variables used in the system. And as always, the way in which those variables are implemented on *different* machines will typically be even *less* likely to have anything other than their higher-level functional properties in common.

This sort of phenomenon is particularly ubiquitous in the social domains. An individual's wealth, a baseball team, a corporation, a story -- all are cases where the identity of token and how it instantiates some higher-level property can vary dramatically in the details of its lower-level implementation over time. The process of change and identity is determined by the rules of the game at the higher rather than the lower level of organization.

In all these kinds of cases, there are really two features that prevent the mapping: One is that *multiple realizability* allows different lower-level structures to implement the same higher-level feature; and the other is that *context-dependence* allows the very same lower-level structure to implement different higher-level objects in different contexts. And context-dependence doesn't allow only for the same properties of some local device contributing differently to the overall functioning of a complex system --- as, e.g., when the same air flow control properties locally could be either functionally a choke or a throttle. Even more commonly, the interplay between levels and context-dependence will be of the sort where higher-level facts about context make the same local part play different functional roles *because they make different properties of that local part functionally salient*. So, consider as an example a switching transistor that produces both a switching DC voltage and also some noisy high-frequency switching transients: The weaker point about context-dependence is just that if the context is different, the transistor switch might --- still taken as a digital switch --- represent different binary functions. But the stronger point is that varying context may make that transistor's switching into a different higher-level event, because the context forces different (lower-level) properties of the transistor to be salient for the overall functioning of the system. Here, context determines what part of the mass of lower-level activity in the object is an implementation of some higher-level function *at all*.

Similarly in the current case: The context of the phenomenological may determine that different aspects of the lower-level activity of the brain become salient in the organism's phenomenology than those which matter to the computational structure of that brain. Different details may be left behind as noise, and different patterns made significant as generalities -- as was the case for the two roles for the transistor. The degree to which computationalist accounts "hide" much of the process and representation from consciousness already indicates that the computational takes as significant various processes and patterns which are not salient in the phenomenology. We have been given no good reason to suppose that the inverse is not true as well.

## 5. Why everybody's wrong

The moral for the idea of a phenomenological-to-computational property identification is clear. Any such suggested account faces the difficulties of the possibility that phenomenological properties are multiply realizable by various computational states that have no particular computational properties in common; the possibility that the very same computational states will underlie different experiences, depending on their position in the more global functioning of the

organism; and --- most importantly, I think --- the possibility that a phenomenological rather than a computational context will make different aspects of the total lower-level swarm of detail functionally significant rather than just bits of noise. With this stack of possibilities running against you, it hardly seems farfetched to suggest that you *might* fail.

This suggestion would then seem to leave Nagel half right, at least. Lycan (and Churchland) fail to rule out the possibility that subjectivity is uncharacterizable except from the perspective of the subject (or one suitably similar, whatever that comes to in the end). It is having that perspective or point of view that allows for our normal ability to refer to those phenomenological properties *as such*; and whether other contexts will allow independent paths for referring to those same properties remains an unanswered question. But even if it turns out that there is no objective characterization of subjective facts or properties as such, and that any such properties *are* importantly perspective-bound, this shouldn't be viewed at all as being anti-*materialist*. Non-reductive identity materialism explicitly *claims* that not all properties are physical properties; that's what distinguishes it from the more reductive accounts.

But this *would* conflict with a functionalist or computationalist view that suggests that subjective mental *properties* (and not just individual mental *states* or *events*) are *identifiable with* (and not just supervenient on) some kind of more "scientifically respectable" properties (e.g. computational or biological properties). This allows for the possibility of a much deeper limitation on functionalist accounts than that suggested by Lycan's "functional" reply to Nagel. His point was just that *knowing* the functionalist account of some subjective state didn't thereby allow you to "know what it's like" to have the state; the "different formats" response seems to me entirely reasonable. But now the problem is that knowing the functional account may not even facilitate *reference* to the phenomenological properties at, in that the phenomenological properties may not be characterizable in functionalist terms --- just as computational properties are not characterizable in physical terms.

So I've argued that Lycan's criticism of Nagel's claim that subjectivity is essentially irreducible doesn't do the job. But I've also suggested that the coherent model that can be provided for essentially irreducible phenomenological properties doesn't impugn materialism in any of its less reductive forms. So in a sense, I'm actually disagreeing with *everybody* about *everything*: On the one hand, Nagel and Lycan both see the possibility of some kind of intrinsically phenomenological properties as being in conflict with materialism. But I've suggested that nothing like anti-materialism follows from such an idea. Intrinsically perspectival (phenomenological) properties don't require anything more than the kind of anomalous monism that motivates the non-reductive conception of function within the materialist framework. And on the other hand, both Nagel and Lycan commit to explicit positions under the current state of the evidence about whether there actually *are* any such intrinsically perspectival properties. In contrast, the position I've argued for takes an explicitly agnostic position on this. Whether there actually are irreducibly perspectival properties or not is for me (unlike for them) an open question which is post-theoretical rather than pre-theoretical in nature.

## **6. Epilogue: Phenomenological themes**

The main thrust of this paper has been to argue that a failure to *reduce* the phenomenological to the functional or computational shouldn't commit you to any kind of anti-materialist view. An identification of phenomenological properties with computational ones is too strong a constraint to place on the kind of pluralistic token materialism assumed by functionalist views in the first

place. Even if phenomenological properties globally supervene on computational ones, they needn't be identifiable with particular computationally characterizable properties. Phenomenological characterizations may well specify coarser-grained properties of the system's activity, exhibiting the kind of multiple realizability and context-dependence with respect to the computational that are the marks of autonomous levels of organization. [See McClamrock (forthcoming), chapter 3]

The possibility that the detection of phenomenological properties might only require agents with the same kind of richness and complexity as ours [I'm not at all assuming that this class couldn't include *artificial* agents; there's no "hollow shell" argument against AI lurking in here.] suggests a possible way to understand the suggestion that there might be fundamentally *human* sciences. If it's the phenomenological taxonomy that's central to some parts of providing an explanation of human thought and behavior, then perhaps the only kind of science of such a domain will use a basic taxonomy which is in *some* constrained sense irreducibly subjective. There might well be explanations of the behavior of complex systems (e.g., human beings in context) where the appropriate taxonomy is such that its properties can only be detected by systems suitably like those being explained; in this case, human cognitive systems.

The idea of intrinsically human sciences is a common one within the phenomenological tradition. But our discussion of the "phenomenological" here so far has been from the standpoint of the philosophy of mind and the philosophy of science. But when in Rome, it's worth at least checking out what the Romans have been up to. So before quitting, let me take note of two further aspects of real phenomenology that may be interestingly relevant here. One is the fundamental distinction in Husserlian phenomenology between *noesis* and *noema*; and the other is the idea that phenomenology might be *existential*, relying on a conception of experience that takes it as inseparable from the existence of the objects on which it's directed. I'll close with a few brief words about each of these in turn. *Noesis and noema*: There is a standard distinction within phenomenology which already makes something much like the contrast I want: Husserl's distinction between *noeses* -- the real temporal parts (or "proper components") of an experience -- and *noema* -- an act's essential *intentional* character (or "correlate"), which is a kind of extension of the notion of Fregean *sinn*. The noema gives a more coarse-grained characterization of the experience than noeses do; the same noema can undergo significant variation over time in terms of the nature of the underlying noetic acts. Noematic features are presented in experience as inhering in the object, and not in the fluctuating state of the subject; they are the "object pole" as opposed to the "subject pole" of experience. So, for example, the noematic color of the paper before me is a uniform white, although part is in shade and part in the light. Such a noematic property "manifests itself in varying perspectives" [Husserl 1964b, p. 261]; it remains the same "whilst the positions of the eyes, the relative orientations, change in many respects, the glance wanders ceaselessly ... and thus in different ways excite the flow of perceptual experience." [Husserl 1964b, p. 261] What unites various conscious noetic acts into a single noema for Husserl is that the various acts make up an intentional directedness toward some particular object or properties in the world perceived as such.

Taking the hint from seeing the noema as intentional correlate rather than proper part of the intentional act, we might similarly in the naturalistic case find the best account of the object-directed phenomenological state not in the unity of some particular representational state, but in the constancy of the object of the experience. The intentional or phenomenological unities in experience -- the taxonomy under which human action has its noticeable systematicities as



rational and goal-directed -- needn't be real, temporal parts of the flow of consciousness, but might be the (abstract, intentional) noematic unities of the real noetic phases of consciousness. The shift in grain between the noetic and the noematic suggests the possibility of a parallel shift between the computational and the phenomenological that leaves the latter as fundamentally non-reducible and autonomous. *The existential turn*: The second phenomenological distinction worth bringing up here is that of the turn toward seeing phenomenology as *existential* --- as relying on a conception of the mind and its states as intrinsically related to a *world*; or perhaps better, as *embedded in its environment*. The central catch-phrase of existentialism -- 'existence precedes essence' -- is meant to capture the general theme that separates existential phenomenology from the earlier "pure" phenomenology of Husserl: that the existence (and perhaps, if you like, *embedding*) of thought and its objects in the real world is in some deep way logically prior to its essence or definition. Questions about the nature and content of thought must be asked while presupposing thought's embedding in the world.

If what defines the unified noema is not a localizable part of the stream of experience, but is instead a coarser-grained structure that's defined intentionally, we then face the possibility that it's not a local or non-context-dependent property of the thinker at all that defines the noematic-grain taxonomy of experience, but instead something about the intentional objects of thought. Even for Husserl, the intentional object of a conscious act is its *real* object in the (external) world. [As Husserl says, "I perceive the thing, the object of nature, the tree there in the garden. That and nothing else is the real object of the perceiving intention. A second imminent tree or even an inner image of the real tree that stands out there before me is nowise given, and to suppose such a thing by way of assumption leads only to absurdity." [Husserl 1964b, p. 243]] If what unifies various noeses into a single noema is that they are directed upon a particular intentional object, then it may be impossible to preserve the intentional properties of the experience while honoring Husserl's methodological constraint of "bracketing out" the external world -- of abstaining from any judgement regarding it, and not making use of the assumption of it.

Similarly, there is now an emerging wave of thought that emphasizes the degree to which the nature of mental activity is intermingled with the world and only analyzable in its interaction with that world. The mind -- even when viewed as a kind of computational system -- is an essentially embedded entity; as such, it is fundamentally misleading to try to analyze it in isolation from the environmental context in which it functions. The understanding of intelligence, thought, and action cannot "bracket off" the structure of the environment, but can only occur and be analyzed in interaction with it. The theme is now coming from a remarkable variety of sources, including the philosophy of language (e.g. Tyler Burge's [1979] 'externalism'), economic psychology (e.g. Herbert Simon's [1981] 'adaptive' 'bounded' rationality), machine vision (e.g. Dana Ballard's [1991] 'animate' vision, and planning in Artificial Intelligence (e.g. Phil Agre and David Chapman's [1987] emphasis on improvisation and interactivity). Ineliminable *interactivity* with the environment is no longer to be seen as an inconvenience for Cartesianism, but is to be embraced and exploited in order to get the psychological trains running on time --- to reduce overhead, simplify representation, and dodge intractability.

The phenomenological (and intentional) view of our mental lives may not then simply redescribe the computational processes in the brain at a different grain of abstraction, but may also fundamentally depend on analyzing them in their contextual embedding in the physical and

social world. Perhaps in such a reconceptualization we can see the final breakdown of the Cartesian illusion of the mind as independent of the physical world for its essence. We "expel things from consciousness and to reestablish its true connection with the world [as] a positional consciousness *of the world*." [Sartre 1966, p. 11]; or as Merleau-Ponty puts it, we see that "the essential point is clearly to grasp the project towards the world that we are." [Merleau-Ponty 1962, p. 405]

[Thanks to Marshall Abrams, Irene Appelbaum, Dan Gilman, Jeff Honnold, and the referees of this journal for their comments on earlier drafts of this paper.]

## References:

- Agre, Philip & Chapman, David (1987), "Pengi: An Implementation of a Theory of Activity", *AAAI 87*, pp. 268-272.
- Ballard, Dana (1991), "Animate Vision", *Artificial Intelligence* 48, pp. 57-86.
- Barwise, Jon & Perry, John (1983), *Situations and Attitudes*, Cambridge, Mass., M.I.T. Press.
- Block, Ned (1978), "Troubles With Functionalism", in N. Block (ed.), *Readings in Philosophy of Psychology, Vol. 1*, Cambridge, Mass., Harvard Press.
- Burge, Tyler (1979), "Individualism and the Mental", in *Midwest Studies in Philosophy*, P. French, T. Uehling, Jr., and H. Wettstein, eds. Minneapolis, Minnesota, University of Minnesota Press, pp. 73-121.
- Churchland, Paul (1984), *Matter and Consciousness*, Cambridge, Mass., M.I.T. Press.
- Davidson, Donald (1978), "Mental Events", in N. Block (ed.), *Readings in Philosophy of Psychology, Vol. 1*, Cambridge, Mass., Harvard Press.
- Fodor, Jerry (1975), *The Language of Thought*, Cambridge, Mass., Harvard Press.
- Husserl, Edmund (1964a), *The Idea of Phenomenology*, translated by W. Alston and G. Nakhnikan, The Hague, Martinus Nijhoff.
- Husserl, Edmund (1964b), *Ideas: General Introduction to Pure Phenomenology*, translated by W. Boyce-Gibson, New York, Collier.
- Leibniz, Gottfried W. (1962), *The Monadology*, in *Leibniz: Basic Writings*, translated by G. Montgomery, La Salle, Illinois, Open Court.
- Lycan, William (1988), *Consciousness*, Cambridge, Mass., M.I.T. Press.
- McClamrock, Ron (1991), "Methodological Individualism Considered as a Constitutive Principle of Scientific Inquiry", *Philosophical Psychology* 4, no. 3.
- McClamrock, Ron (forthcoming), *Existential Cognition: Computational Minds in the World*.
- Merleau-Ponty, Maurice (1962), *The Phenomenology of Perception*, London, Routledge & Kegan Paul.

Nagel, Thomas (1979), "What is it Like to Be a Bat?" in T. Nagel, *Mortal Questions*, Cambridge, Cambridge University Press, pp. 165-180.

Sartre, Jean-Paul (1966), *Being and Nothingness*, translated by H. Barnes, New York, Washington Square Press.

Sartre, Jean-Paul (1957), *The Transcendence of the Ego: An Existentialist Theory of Consciousness*, translated by F. Williams and R. Kirkpatrick, New York, Noonday Press.

Searle, John (1980), "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3, pp. 417-24.

Simon, Herbert (1981), "The Psychology of Thinking: Embedding Artifice in Nature", in H. Simon, *The Sciences of the Artificial*, Cambridge, Mass., M.I.T. Press, pp. 62-98.