Burge's Dualism

Bernard W. Kobes
Arizona State University

draft: 18 May, 2008

Slated for *The Waning of Materialism*, George Bealer and Robert C. Koons, eds. (forthcoming from Oxford University Press).

Introduction.

In his 1999 Presidential Address to the Pacific Division of the American Philosophical Association, Tyler Burge contrasted the inflated conceptions of both philosophy and humankind at the ends of the 17th, 18th, and 19th centuries— for example in the works of Leibniz, Fichte, and Bradley— with various twentieth century philosophical deflationisms. A task for philosophy, Burge proposed, is to articulate the specialness of humankind while avoiding the inflations of the past. For example, anti-descriptivism in philosophy of language, anti-individualism in philosophy of mind, and externalism in epistemology, exhibit 'how our natures are determined by norms that reach beyond what we as individuals control. We can better understand the ways that rational beings depend on a universe that is not made up of structures of reason at all.'[1]

An instance of this larger theme is the relation between Burge's anti-individualism about representational aspects of mentality and his proclivity for a

weakly dualist position on the mind-body problem. My title, 'Burge's Dualism', is perhaps hyperbolic, as one does not find in Burge's writings to date any unequivocal affirmation of mind-body dualism. One does find, however, a pervasive set of methodological stances, attitudes, and *dicta* congenial to dualism. One also finds a sharply delineated argument against the physicalist token identity theory, a set of reflections on mental causation that debunk the alleged problem of how mental properties can be causally relevant in a fundamentally physical cosmos, and a more diffuse argument against a weaker materialism that takes the mental to be composed of the physical. For these reasons I will continue to refer to 'Burge's dualism', while acknowledging, in the end, a certain diffidence in his view of the matter.

After gathering Burge's dualist themes and anti-materialist arguments, I will present some prima facie resistance to Burge's doubts about material composition, and explore a way in which his stance might be developed to meet such resistance. I will not attempt a full-dress evaluation of Burge's dualism, which in any case would be premature. My goals are to organize the elements of Burge's dualist stance, to convey a sense of their depth and seriousness, and to gauge what further articulation and empirical confirmation would be needed to set the view among prominently available options for responding to the mind-body problem.

Burge's methodological orientation.

Burge's dualism does not focus on phenomenal properties of conscious experience; he does not, for example, appeal to Cartesian or Kripkean intuitions about the conceivability of the physical and the phenomenal coming apart, or to the alleged explanatory gap between them, or to not knowing phenomenal facts despite knowing all pertinent physical facts. Burge's dualism, at least in his published writings to date, concerns intentional or representational aspect of mentality. It may be, however, that some of the considerations he brings to bear on intentional aspects will have analogues for phenomenal aspects of experience.[2]

One manifestation of twentieth century deflationism is the widespread presumption, in the latter half of the century, of materialism as a default— the presumption that we ought to be physicalists if we can— and a preoccupation with physicalist sketches of mentality. Burge rejects the presumption of physicalism, which he sees as a pervasive ideology without clear foundation in either compelling apriori metaphysics or in successful explanatory practices.[3] Of Jaegwon Kim's assumption, typical of much contemporary philosophy, that the world is 'fundamentally physical', Burge complains that the outlook is under-specified: 'There are many questions to be raised about this idea and how it is supposed to apply to various cases (the mathematical 'world', the 'worlds' of value, right and

wrong, beauty, rational justification, semantics, indeed mind).'[4] 'It is not obvious to me', he writes in discussing Descartes, 'that it is mistaken to suppose that mental agents and their mental powers, acts, and states are in no literal sense physical.'[5]

Burge's dualism is naturalistic, in a weak sense of that word:

> The theme in naturalism that deserves the status of orthodoxy is not its materialism and not its demand that mentalistic discourse be given some ideologically acceptable underpinning. It is its implicit insistence that one not countenance any form of explanation that will not stand the scrutiny of scientific and other well-established, pragmatically fruitful methods of communal check and testing. (More crudely put, it is the opposition to miracles and to postulation of unverified interruptions in chains of causation among physical events.) [6]

The theme of epistemic reliance on our successful explanatory practices is fundamental and pervasive. Epistemic strength in philosophy derives, Burge insists, from our successful commonsense and scientific explanatory practices. Burge does affirm the possibility of substantial apriori metaphysical knowledge of the nature of the mental. But he holds that metaphysics ought to be pursued in an exploratory spirit, and with a sense of its own fallibility. Metaphysical principle is trumped by consideration of the nature and needs of successful explanatory practice.

The fullest expression of Burge's methodological pragmatism occurs in "Individualism and Psychology":

> Not just questions of supervenience, but questions of ontology, reduction, and causation generally, are epistemically posterior to questions about the success

> of explanatory and descriptive practices. One cannot reasonably criticize a purported explanatory or descriptive practice primarily by appeal to some prior conception of what a 'good entity' is, or of what individuation or reference should be like, or of what the overall structure of science (or knowledge) should turn out to look like. Questions of what exists, how things are individuated, and what reduces to what, are questions that arise by reference to going explanatory and descriptive practices.[7]

Burge accepts, at least for the sake of argument, a global supervenience of mental facts upon the totality of physical facts.[8] But from this, little of ontological or causal interest follows. Successful explanatory and descriptive practices occur at a variety of levels that, as a matter of empirical fact, proliferate and only occasionally reduce. This entrains a proliferation in our attributions of cause and effect and in our presumed ontology. 'The world is a rich and complicated place.'[9]

Anti-individualism.

Burge has famously argued that mental states and events are individuated by broad features of the subject's physical, social, or historical environment. Mental states are not individuated purely by their physical, functional, or phenomenal features, insofar as these can be fully characterized apart from the subject's environment. This view is often described as anti-individualism about mental content, but the point directly concerns the natures of mental states and events as such, not merely their referents, and not merely their contents considered as *abstracta*. According to the view, there are deep individuative relations between

features of a subject's environment and her representational mental states, considered in their representational aspect. Not only the referent, but also the *way* in which the referent is thought of or perceived, is environmentally individuated.

Individuation is typically mediated by (perhaps indirect) non-intentional, causal relations to things in the subject's environment. Such causal, non-intentional relations to represented things may be further mediated by the subject's linguistic community, or by the subject's species ancestors. Though causally mediated, individuation is not itself a form of causation, but rather conceptual or metaphysical determination of identity. The laws by which mental states are caused by environmental variables are not at issue here, still less 'nature versus nurture'. Individuation by reference to the environment can give rise to, and explain, failure of local supervenience. That is, neurological, narrow functional, and narrow phenomenal states may be held constant across worlds, while representational states vary in virtue of conceived environmental variation. But anti-individualism is a doctrine of how mental states are individuated, hence a doctrine of their nature, and not fundamentally about the role of a community, deference to experts, or failure of local supervenience.

Burge has developed and defended this view by means of a series of detailed and powerful thought experiments. The thought experiments come in four varieties:

those that feature natural kind concepts, those that key on reliance on a linguistic community to fill out incompletely understood concepts, those that feature perceptual contents, and those that highlight the possibility of challenging even fully understood meaning-giving explications. Space limitations prevent detailed exposition of the thought experiments, but I will say a word about each type.

The first type of thought experiment resembles Hilary Putnam's famous Twin Earth thought experiments about the meanings of natural kind terms. In Burge's version, a subject on Earth thinks a thought about a natural kind, for example that aluminum is a lightweight metal. We imagine that on Twin Earth a duplicate of the subject exists who is identical to her in all physical, functional, and phenomenal respects, insofar as these can be non-intentionally characterized. On Twin Earth, however, there is no aluminum; instead, there is a distinct metal that is similar to aluminum in superficial respects but that differs from aluminum in microstructure. The microstructural differences make for macro-level differences that could be detected by scientists, but these happen not to causally impact the subject differently from her twin. It is plausible that the twin does not think that aluminum is a light-weight metal. The subject and her twin think different thought types, despite being exactly similar in all individualistic respects. The difference, Burge argues, affects the representational character of thought; it is not merely a difference in reference,

and in particular cannot be understood as a difference in indexical reference.

In the second type of thought experiment there are no differences in the environmental natural kinds on Earth as compared to Twin Earth. Instead, the subject thinks that he has arthritis in his thigh, but does not know that the word 'arthritis' correctly applies only to certain rheumatoid ailments of the joints. On Twin Earth, the subject is duplicated in all individualistic respects, but in that linguistic community the word 'arthritis' is standardly applied to a wider class of rheumatoid ailments, including the one that the subject and his twin have in their thighs. It is plausible that the Twin Earth subject does not think that he has arthritis in his thigh. So the subject and his twin think different thoughts, despite being exactly similar in all individualistic respects. The representational difference in this case derives from the subject's deference to communal experts for the correct application of his concept, which he thinks with but has not completely mastered.

In the third type of thought experiment, Burge imagines a creature that is visually attuned to cracks, while its twin is visually attuned to shadows. The physical, functional, and phenomenal states of the creatures, insofar as these can be individualistically described, are imagined to be exactly similar. Burge argues that there is a difference in perceptual representation— how the creatures visually represent the world to be— and not merely a difference in the objects seen. The

creature on Earth perceives cracks as cracks; its twin perceives shadows as shadows. Individuation of representational states is again mediated by non-intentional, causal relations to things in the environment, though in primitive perceptual cases the mediating causal relations may be borne by the creature's species ancestors.

Burge is not committed to the nomological possibility of failure of local supervenience in perceptual cases. For it may be that any environmental difference that makes an individuative difference to perceptual events also makes a correlative difference to neurological processing; thus local supervenience may be underwritten by the tight responsiveness of perceptual neural events to environmental variation. Even so, the thought experiments suggest that environmental differences directly, and not (and not *via*— hence 'directly') the neural differences, *best explain* the differences in individuation of perceptual states. In any case, vision science, especially in the research tradition emanating from David Marr, assumes or presupposes perceptual representational contents that are environmentally individuated, or so Burge argues.

In the fourth type of thought experiment, the subject has mastered communally agreed criteria for the proper application of a word or concept. He knows, for example, that competent speakers, including experts, take sofas to be

pieces of furniture of such-and-such construction made or meant for sitting. But the subject doubts that these explications are correct, and concocts an alternative theory: the nature of sofas is to be works of art or religious artifacts of a certain sort. The subject's theory is false, but his twin's corresponding theory is correct: the nature of things called 'sofas' on Twin Earth *is* to be works of art or religious artifacts. Burge argues that the subject and his twin think thoughts of different mental types, despite being similar in all non-intentional respects, and despite knowing what are the communally agreed meaning-giving explications for 'sofa'. This thought experiment shows that incomplete mastery of concepts and deference to experts, which figured in the arthritis thought experiment, are not necessary to exhibit the anti-individualistic phenomena. Meaning-giving explications can be challenged, as they are in the 'sofa' thought experiment, and it can be questioned whether communal patterns of use are as they should be. Mental states are individuated so as to allow for the possibility of such challenges. Meaning-giving explications are empirically substantial; they concern an external, objective subject matter, to which thinkers have independent, causally mediated access.

The four types of thought experiment variously manifest a unitary, deep feature of mental representation, namely, that we represent an objective subject matter, things whose natures are public and independent of any creature's mental

events or representational acts. An objective, mind-independent world can be mentally represented only if the relevant mental states derive their natures in part from the natures of things represented. This derivation of natures is not systematically mirrored in the nature of some local neurological or functional substrate, nor in patterns of individual or communal use, nor in conceptual or linguistic mastery. Instead, there is a 'cognitive distance' between thinker and represented objects, so that mental individuation is directly mediated by non-representational relations between perceiver or thinker and represented objects.

Against token identity.

Anti-individualism, so understood, seems incompatible with the strongest kind of mind-body dualism, which employs the traditional conception of a substance as something capable of existing all by itself, without any other contingent thing of an equally basic ontological category. On the other hand, twentieth century materialism militated against recognition of anti-individualist factors in the individuation of mental states; conversely, anti-individualism poses obstacles to certain forms of materialism. Neural states and events are individuated relatively independently of the subject's environment. A neural state, whether type or token, does not partly derive its nature from the natures of things which the subject interacts with and represents. So if a mental state is environmentally individuated in

the ways that Burge has argued it is, and if this is not merely a contingent feature of the state but part of its nature, then the relation between the mental state and any underlying neural state coincident with it cannot be identity.

The point is sharpest when local supervenience fails. Subject and twin can be imagined to share their neural states and events, even as their thought contents vary with imagined variation in the environment. A mental state is identified by its intentional content; having the content it has helps make it the mental state it is. The neural state could exist without the corresponding representational mental state. The mental state and the neural state have distinct modal profiles.

A similar modal argument seems to show that mental states cannot be identical to functional states, if functional states are identified narrowly, with the inputs being sensory stimulation and the outputs being muscular activity or bodily motion. For narrow functional states can stay constant across twins, even as mental kinds vary. A broader kind of functionalism that identifies inputs and outputs in terms of environmental kinds, and is also socially and linguistically distributed, might escape the most straightforward objection of this form, but such 'long-arm' functionalist proposals are excessively programmatic.[10]

A weak and widely held form of materialism identifies each mental token, each mental event in a subject on an occasion, with some neural event in the subject

on that occasion. The 'token identity theory' is typically combined with a denial that mental types can be identified, or even universally correlated, with neural types. Burge argues, however, that a representational thought token cannot be identical to any neural token. For any plausibly relevant token neural event could occur in a 'twin' environment, construed here as a distinct possible world, such that the content of the thought token in that world differs from the content of the actual thought token, in virtue of environmental differences. The neural event could be the same token across worlds, even as the mental contents differ. In our descriptive and explanatory practices we commonly identify thought tokens by such basic factors as the relevant subject, time, and representational content. By contrast with, say, sentential forms, we have no other standard way to identify thoughts; in particular, we have no way to identify a thought so that it has its content only contingently, and there is no reason to think that a future cognitive science will do so. So it is plausible that no thought token could have had a different content from the one it has. Indeed, Burge suggests that this is evident and apriori, and that to deny it would amount to changing the subject.[11]

I said that according to Burge, anti-individualism is fundamentally a thesis about the individuation of mental states, and not about failure of local supervenience. Anti-individualism is a precondition of mental representation of an

objective, mind-independent world. The thought experiments display various kinds of cognitive distance between subjects and represented things. The cognitive distance phenomena show that mental states derive their natures in part from the natures of objective, mind-independent objects to which the subject stands in (perhaps highly indirect) causal, non-intentional relations. The same goes for tokens: a thought token's intentional content is part of its nature, so it derives its nature in part from the natures of objective, mind-independent external objects. But it is implausible to suppose that any neural token derives *its* nature in anything like this way. Our most basic ways of identifying neural tokens are through the descriptive and explanatory practices of neuroscience. This amounts to an anti-token-identity argument that does not exploit failure of local supervenience or distinctness of modal profiles. I am not aware of Burge having argued anywhere against token-identity in precisely this manner. But it seems to me that, for systematic reasons, he ought to be willing to accept this non-modal argument. This argument casts its dialectical net less widely, since it depends on a more theoretically embedded and perhaps less evident metaphysics; fewer may find its premises compelling. On the other hand, the argument may be more illuminating, supplying a deeper explanation of why mental and neural tokens are distinct.

Mental causation.

Mental causation of physical events, and physical causation of mental events, have been seen as posing difficulties for dualism at least since Descartes. How could two such ostensibly different realms interact causally? Identifying mental events with physical events has often been thought to help solve puzzles of causal interaction. If mental event tokens are identical to physical event tokens, then causal interaction involving mental event tokens is just a species of causal interaction involving physical event tokens.

On the other hand, if mental properties are not reducible to physical properties, as token-identity theorists typically hold, then it can seem puzzling how mental properties can be causally relevant. If the causal efficacy of a mental token is due to its physical nature, then it can seem that none of its causal efficacy is due to its mental nature. By such reasoning mental properties can come to seem epiphenomenal— merely along for the ride. Compare the manifestation of a phenotype in a biological parent; this feature does no causal work in effecting the inheritance of the characteristic by offspring. The real causation goes on at the level of genes. The agenda for much recent writing on mental causation has been set by the view that mental causation is underwritten by the physical natures of mental tokens, and that if mental properties are not reducible to physical properties, then

their causal relevance is indirect at best.

Burge holds that these worries about epiphenomenalism show a loss of perspective, a set of misplaced epistemic priorities. What we know about mental causation derives primarily from mentalistic explanatory practices, commonsense and scientific. Psychology employs an intentional idiom, deepening and refining its explanations of thought and action in representational terms. This is an adequately robust scientific enterprise, with no serious sign that it will ever be supplanted by any non-intentional explanatory practice, and its explanations are causal on their face. We can be justifiably confident, therefore, that mental causes can have mental effects, and sometimes physical effects, and that mental properties are causally relevant. This is also a presupposition of our status as agents, and a precondition of our rational deliberations having any point. Mental causation does not stand in need of being underwritten by the physical nature of mental tokens.

A variety of metaphysical principles have been employed to cast doubt on the possibility of irreducible mental properties being causally relevant: that, for example, the causal ancestry and causal posterity of a physical event can consist only in other physical events physically described, or that there is a tension between a mental event M's being instantiated *because of* its physical supervenience base, and M's being instantiated *because of* some prior mental event. Kim has argued

that such principles pose a credible threat to the causal relevance of mental properties, and has argued further that in responding to the threat we learn that token physicalism is not physicalism enough. We need a reduction of mental properties ultimately to physical properties; mental properties are causally relevant, on Kim's view, but only insofar as they reduce to physical properties.

Burge responds that the appearance of a threat to the causal relevance of mental properties is bogus. What we know from explanatory practice is far more secure than any metaphysical principle that has been employed to conjure up the appearance of such a threat. Burge also subjects the metaphysical principles to specific, searching criticism. Some fail to adequately distinguish distinct levels of explanation and cause, which may coincide on an effect without any objectionable or coincidental kind of over-determination. Some fail due to illicitly treating mental causation as supplying some extra bump or energy that would interfere with physical causation. Some fail to adequately distinguish causation from the kind of determination that characterizes the supervenience relation. Kim holds, for example, that there is a tension between 'horizontal determination', a mental event M's causing a later mental event M*, and 'vertical determination', M*'s synchronic dependence on its neural supervenience base P*. Kim's resolution of the tension proceeds by way of the principle that M causes M* by way of M's supervenience

base P causing M*'s supervenience base P*. But given anti-individualism this resolution of the alleged tension is not credible, Burge argues, for a belief or thought will typically have an intractably complex supervenience base, spread over large expanses of space and time.[12]

The upshot, according to Burge, is that materialist theories of mind have failed to illuminate mental causation, despite that being advertised as a key point in their favor. Important metaphysical questions about mental causation remain open. Burge acknowledges that physics is gapless, that physical causes do not leave openings for mental causes to fill. The physical asymmetrically sustains the mental, and global supervenience is presumably true. But these weakly naturalistic points show little about the nature of mental-mental or mental-physical causation, and are in any case poorly understood. Mental and physical causes are not basically 'the same', but neither are they 'in tension'. How mental causation is to be understood in relation to physical causation is a real issue. Mental and physical causes operate systematically in concert, are not in competition, and do not trade in massive coincidences. A metaphysical understanding of how this occurs, from a perspective that encompasses multiple levels simultaneously, is to be desired.

Doubts about material constitution.

A still weaker form of materialism than token identity holds that mental states

and events are *constituted* or *composed* of physical— presumably neural— states and events. Compositional materialism seems to escape Burge's argument against token identity, because the same neural complex may compose one mental token in the actual world, and a distinct mental token in the twin world. Since composition is not identity, a mental event token may derive its nature in part from the natures of represented things in the environment, while the neural event token that composes it does not. Perhaps compositional materialism about the mental will help us understand how mental and physical causes can operate systematically in concert, without competition or coincidence. Presumably too it will help us understand the asymmetrical dependence of the mental on the physical. Certainly we can manipulate the mental by manipulating the neural in an intricate variety of healing, useful, recreational, nefarious, or tragic ways. This too is a matter of a successful scientific explanatory enterprise.

In several places, however, Burge expresses doubt about compositional materialism. He notes that the paradigm of decomposition of material objects into physical particles does not apply straightforwardly to properties or events, let alone to numbers, intentional contents, and methods.[13] We do not identify thoughts by their physical compositions or constitutions. The Battle of Hastings, and the emergence of North America as a continent, resemble thoughts in being non-

individualistically individuated. But they are plainly physical and are typically identified by their physical parts, and in this respect they contrast with mental events.[14] Some physical events, such as particular 'wars, avalanches, thunderstorms, meal cookings' may not fall under the kinds of any natural science, yet they, unlike psychological states, explicitly and obviously involve ordinary physical properties that are used in explanations in the physical sciences.[15]

Material constitution is a scientific notion with specific explanatory uses in, for example, chemistry and physiology. We explain how chemical and physiological kinds interact by reference to their physical constitutions. How molecules interact with each other, and how atoms interact with each other, is explained by how their respective component parts interact.[16] But actual successful mentalistic explanations, commonsense or scientific, do not appear, on their face, to refer to material compositions of mental states.[17] Moreover, intentional mental states often serve as explanantia, not merely as explananda; they enter substantially into explanations, with no reference made to their putative material parts.[18]

Burge understands Descartes to hold that the individuation of particular minds is primitive: a mind is an agent of particular mental acts. 'I believe that Descartes may be on to something important in regarding thinkers as consisting not in some special sort of stuff, but in particular instances of the special type of agency, power,

consciousness, and point of view involved in thinking.'[19] This suggests a view on which the fundamental sortals in psychology will key on activity, not constitution. The view marks a distinction between psychology and material sciences such as chemistry, physiology, and geology, in which constitution sortals play a fundamental explanatory role. The manipulation of neural events to effect mental changes, and the study of neural activity through imaging technologies such as fMRI, aim to specify the underlying neural events that sustain thought— where thought is conceived as the activity of a mental agent. But it would be a mistake, on the Cartesian conception to which Burge is tempted, to think of such underlying events as what thoughts or mental agents *consist in*. Our most fundamental explanations of mental activity allude to factors such as having reasons, and not explicitly to the matter which thoughts ostensibly comprise, in the manner of chemistry and other sciences of matter.

There is an epistemic possibility that we will some day make explanatory use of material constitutions of mental states, but Burge sees no positive reason to expect that this will be a feature of a fully developed psychology. To expand our sense of the relevant possibilities, he writes:

> Maybe science will never make use of anything more than limited correlations with the lower, more automatic parts of the cognitive system. Maybe identities or part-whole relations will never have systematic use. Maybe the

> traditional idea of a category difference will maintain a presence in scientific practice.[20]

These remarks convey something of the open, questioning, exploratory nature of Burge's doubts about physical composition of the intentional.

It can seem, however, that these doubts leave us with an enervated conception of the explanatory goals of cognitive neuroscience. It remains less than satisfyingly clear how material composition of intentional events could turn out empirically false, assuming that more basic events can compose or constitute higher-level events. Composition is, after all, even weaker than token identity, and the considerations Burge adduces turn on intentionality, not the peculiarities of phenomenal character. Surely some principle of theoretical simplicity or unification favors material composition, even if full-fledged ontological reduction is frustrated by the argument against token identity. There is no shortage of events going on in the brain to do the constituting work, and if our actual future science never makes use of intentional-neural correlations, that may reflect some merely practical limitation. In response to these points of prima facie resistance I will sketch a metaphysical picture which, if it turns out to be empirically accurate, would vindicate, I suggest, Burge's doubts about physical composition.

How material composition of the intentional could be empirically false.

Doubts about material composition can be seen as an effort to free our metaphysical and empirical imaginations, to open us up to alternative possibilities in a philosophical climate in which it is difficult to see how cognitive neuroscience could fail to support, at the very least, material composition of intentional tokens. To help make the possibility vivid I invoke a 'mathematical archangel' (C. D. Broad's term), a being with unbounded logical and mathematical abilities, who begins with only the complete facts about fundamental physical objects, events, fields, laws, and causes over all of space and time. This idealized calculator is a (dispensable) heuristic device to depict a conjectured order of metaphysical dependence. The archangel can compute the chemical and physiological facts, I suppose, up to and including the biological and neural facts over a community of psychological subjects. Assuming global supervenience of the intentional psychological on the physical, and skirting issues of phenomenal character, I further suppose that he can compute the intentional facts over the community of subjects.

The crux is whether in doing so the archangel has available a 'compositional short-cut' at the neural level, that is, whether he could in principle discern at the neural level, prior to attributing propositional attitudes to the subjects, a network of states and events that interact causally, and that are candidates for composing or

constituting the intentional states and events to be attributed at a later stage of the computation. Having first identified these discrete neural states and events, the archangel could later see them as constituting intentional mental states and events, identifying the latter by reference to environmental kinds, in anti-individualistic fashion. The actual progress of cognitive neuroscience may support this metaphysical picture— but then again, it may not. It may instead support an alternative picture in which the archangel must first reconstruct the intentional psychology of his subjects, attributing propositional attitudes in something like the way we attribute them in commonsense discourse and in intentional psychology. On this second picture, not even an idealized calculator with perfect knowledge of the basic physical facts could compute, in advance of intentional psychology, particular discrete neural events that compose or constitute particular intentional events.

The suggestion here is not that any sort of mental fact would be forever closed to the archangel, but only that we can understand failure of material composition as the unavailability to the archangel of any prior neural identification of intentional events. Compositional materialism is false just in case the archangel would have to first recapitulate our interpretive practices and intentional psychology, and only then, if at all, seek neural event correlations. The archangel may have initially available no 'handle' on a token intentional event other than such

features as its subject, time, and representational content. The intentional theory the archangel constructs may be identical to, or close to, ours. Perhaps, indeed, the archangel must attribute intentional states to the subjects, and simultaneously attribute a systematic practice of intentional *attribution* to those subjects; the intentional events themselves, and subjects' practices of attributing and self-attributing them, may be metaphysically intertwined. But this latter hypothesis is ancillary to the main proposal. If token neural correlates can be found at all for token mental events, the direction of metaphysical explanation is from mental events and their causal patterns to the correlated neural events. By contrast, the archangel first identified chemical and physiological kinds and their instances by their physical components, distinctively arranged.

Indeterminacy of the sought-for neural tokens may frustrate the archangel's search for principled local correlations for particular thoughts; he may not even find plausible candidates for material constitutions of particular thoughts. Given anti-individualism, a thought event *m* will have no minimal supervenience base *n* within the brain— no smallest neural event *n* such that *m* will occur in any possible world in which *n* occurs. It would be a mistake, moreover, to think of thought tokens as in part environmentally constituted, as smeared out over the world. The individuating factors may involve vast tracts of space and time, without discrete or natural

boundaries. Individuation may be mediated by causal factors, such as patterns of deference within a community, distinct from the individual subject's thought. The suggestion of 'smeared constitution' appears to conflate individuation and reference; recall that anti-individualism concerns not only reference, but also the way in which the referent is represented.

Some principle would therefore be required to demarcate the boundaries of the inner neural event that composes a particular thought. But we have no advance guarantee that such determinate boundaries exist. Thoughts often seem to involve mental agency, and mental agency may not be explicable in terms of causal interactions among sentences in a neurally instantiated language of thought. If thoughts are mental acts, it may be unclear how much of the agent to include in the neural composition of the thought. Moreover, if a thought includes the exercise of constituent concepts, then the thought's neural constitution would presumably include neural proper parts corresponding to the exercise of those concepts. But the archangel may lack any principled way of demarcating the boundaries of neural events whose neural part-whole relations respect the thought's conceptual structure. The relation between a thought and its constituent concepts is distinct from the relation between the thought and its putative neural composition, or the relation between an underlying neural event and its neural components. It is an open

empirical question whether these can be brought into non-arbitrary alignment.

In any case anti-individualism helps dispel any lingering suggestion that a thought *must* derive its identity and determinacy from its neural constitution. For if an intentional mental event inherits its nature in part from environmental things it is about, then that is an independent source of the event's identity and determinacy. (Here it is worth noting that, while the Kripkean judgment that *this table* could not have been made from a different block of wood has considerable pre-theoretic appeal, the judgment that *this thought* could not have been subserved by a different neural event lacks equivalent pre-theoretic appeal.[21]) The thought is the event that it is, partly because of its having inherited its nature from things it represents, and does not *stand in need* of material constitution to underwrite its status as an entity. An eliminativist strain in American philosophy of mind since Quine holds intentionality hostage to the demand that it stand in a properly disciplined relation to the behavioral or neural realms. Burge's dualism consists partly in his preparedness to jettison even the weakest such demand— material constitution— if it does no explanatory work of a certain kind, and to let intentional explanations stand on their own.

The distinctive kind of explanatory work that, according to Burge, is done by the notion of composition gives rise to an independent doubt about material

composition of intentional events, a doubt that could persist even if we assume that the archangel does, in the end, compute principled and determinate correlations between intentional and neural events. Recall the point that we explain the causal interactions of chemical and physiological kinds in terms of their material composition, and similarly for geological, astronomical, and neural kinds. We do not, however, explain the causal interactions of intentional states in terms of their material composition, at least not yet. We can understand the causal necessity of a chemical interaction over time as deriving from the causal necessity of interactions at the molecular and atomic levels. But mental states are often brought in as explanantia, without explicit reference to underlying neural events. Would the archangel's intentional-neural correlations do any explanatory work analogous to that of composition in familiar sciences of materially constituted things?

This is an open empirical question. The actual progress of cognitive neuroscience may support the following metaphysical picture. The archangel's intentional-neural correlations yield an account of how neural events asymmetrically and synchronically sustain certain intentional mental events, but do not illuminate diachronic causal relations among them, at least not in the manner of the familiar sciences of materially constituted things.

For example, suppose that neural event tokens n1 and n2 underlie intentional

mental event tokens M1 and M2 respectively, and that in some natural explanatory context it would be correct to say that M1 causes M2. It may be, however, that n1 does not appear to the archangel as, in any illuminating sense, *the cause* of n2. It is compatible with this to suppose that n1 is counterfactually relevant to n2, in the dime-a-dozen sense that, had n1 not occurred, n2 would not have occurred. Perhaps n1 is among a plethora of neural events that are *causally relevant* to n2, without the archangel being able to enter into an explanatory context in which he sees n1 as *the cause* of n2, prior to his construction of intentional psychology. In a derivative and retrospective sense, the archangel may later see n1 as *the cause* of n2, in light of the previously computed causal relation between M1 and M2. But if n1 had been, in any illuminating sense, *the cause* of n2, then the archangel would have been able to compute that fact prior to computing the subject's intentional psychology, and *ex hypothesi* no such computation was available to him. Only after M1 and M2 had been anti-individualistically computed did the relation between n1 and n2 come to the fore. I suggest that it is epistemically possible that this metaphysical picture will be empirically supported, and further that if it is supported, then the explanatory role of the underlying neural events is quite unlike that of component physical entities and events in familiar sciences of materially constituted things.

What is distinctive in this metaphysical picture is that although specific neural

events asymmetrically and synchronically sustain corresponding intentional mental events, intentional causal relations stand on their own, without needing or deriving support from causal relations at the neural level. The necessity that attaches to intentional mental causal transactions is *sui generis*, and need not derive from the necessity that attaches to neural causal transactions. The notion of constitution is conceptually linked to distinctive kinds of diachronic causal explanation, so we have depicted a world in which the intentional is not materially constituted. Again, the heuristic device of the mathematical archangel is meant to vividly depict certain possible patterns of metaphysical dependence. The depicted metaphysical dependencies are inspired by, and expand on, Burge's expressed doubts about whether intentional mental events are materially constituted. Whether these possible metaphysical dependencies are actual is an open empirical question.

The intentional in a physical cosmos.

The metaphysical picture here sketched conflicts with a physicalist account of the explanatory goals of cognitive neuroscience, and how that discipline illumines the nature of thought and intentional causation. But the worry that it leaves us with an enervated conception of the deepest explanatory goals of cognitive neuroscience is ill-founded. We may hope for a richer understanding of the asymmetric synchronic sustenance of the intentional by the neural; in this sense we gain an

understanding of how our mental lives are *possible*. We may particularly hope for a more articulated understanding of the *compatibility* of the intentional and the neural, how it is that they do not pull apart from each other. For the question why they do not pull apart can retain its grip on us even after we acknowledge that the necessity of intentional causation does not derive from that of physical causation, and does depend on physical notions such as energy transfer. Neural events sustain mental events without suffering interference. The physical goes on in the way that it will, without coincidences or inexplicable miraculous parallels to mental processes. We may see cognitive neuroscience as fleshing out in satisfying detail these skeletal points, even if the science turns out not to illumine the identity or constitution of mental events, or the nature of mental causation.

Note that substantial chunks of the physicalist world picture are not being called into question here. A Burgean dualist could hold that (a) there is no old-fashioned mental substance capable of existing by itself, no soul in the sense of traditional metaphysics, no ectoplasm; (b) the intentional globally supervenes on the physical; (c) the physical is causally gapless at its own level, modulo quantum indeterminacy; and (d) intentional mental events are synchronically and asymmetrically sustained by underlying neural events. An appropriately qualified desideratum of theoretical simplicity in physics and metaphysics may be satisfied by

these points. Of course our metaphysics must not depict the world as simpler than it is, given evidence that it is a 'rich and complicated place'.

It will be evident that Burge's position on the mind-body problem is a quite weak form of dualism. Combined with the four theses of the previous paragraph, it could aptly be described as an extremely weak form of physicalism, in contrast with inflated dualisms of past centuries. Perhaps that accounts for the note of diffidence in Burge's position when he writes,

> It seems to me that philosophers should be more relaxed about whether or not some form of materialism is true. I think it a thoroughly open— and not very momentous— question whether there is any point in insisting that mental events are, in any clear sense, physical. ... What matters is that our mentalistic explanations work and that they do not conflict with our physicalistic explanations. But it serves no purpose to over-dramatize the conflict between different ontological approaches ... [22]

Of course the particular arguments and positions on individuation, identity, causation, and constitution are what matter. There is nevertheless a certain dramatic sweep to the larger conception in which these philosophical particulars are set. According to that larger conception we can limn metaphysical accounts of how irreducibly intentional events and causal processes might arise and persist within a fundamentally non-intentional physical cosmos. These accounts interact with our empirical theories. We thereby describe ways in which 'rational beings depend on a universe that is not made up of structures of reason at all.'

Notes

I wish to thank Torin Alter, David Braun, Derk Pereboom, and Steven Reynolds for helpful discussion of these matters.

Most of the relevant papers are collected in Tyler Burge, *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford University Press, 2007).

1. Tyler Burge, "A Century of Deflation and a Moment About Self-Knowledge", *Proceedings and Addresses of the American Philosophical Association*, Vol. 73, No. 2 (November 1999), pp. 25-46. The quoted material is from p. 30.

2. Burge has written two articles that include substantial discussion of phenomenal consciousness: "Two Kinds of Consciousness", *Foundations of Mind*, Chapter 17, and "Reflections on Two Kinds of Consciousness", *Foundations of Mind*, Chapter 18. Both articles take phenomenal consciousness to be the fundamental notion, but also articulate a distinct notion of rational-access consciousness. On p. 418 of the latter article, Burge rejects the supposition that understanding consciousness will require solving the 'hard problem', so that consciousness will remain a cosmic mystery unless some sort of functional or neural reduction of the phenomenal is achieved. We do not yet know how to non-reductively systematize and integrate phenomenal consciousness with respect to empirical science. But we are beginning to understand mental representation this way, and we have no reason to doubt that this will be possible for phenomenal consciousness.

3. See *Foundations of Mind*, p. 360, for strong statements along these lines.

4. Postscript to "Mind-Body Causation and Explanatory Practice", *Foundations of Mind* Chapter 16, p. 368.

5. "Descartes on Anti-individualism", *Foundations of Mind*, p. 434.

6. "Philosophy of Mind: 1950-2000", in *Foundations of Mind*, p. 447.

7. Burge, "Individualism and Psychology", ... Burge traces this methodological pragmatism, insofar as it touches ontology, to Frege and to Quine.

8. *Foundations of Mind*, pp. 369-370.

9. Cf. *Foundations of Mind*, p. 348. The quoted line, taken harmlessly out of context, is from p. 28.

10. On the excessively programmatic nature of an externalized functionalism, see *Foundations of Mind*, p. 454.

11. *Foundations of Mind*, p. 157.

12. *Foundations of Mind*, p. 374.

13. *Foundations of Mind*, p. 357, note 14.

14. *Foundations of Mind*, pp. 229-230.

15. *Foundations of Mind*, p. 357.

16. *Foundations of Mind*, p. 359.

17. *Foundations of Mind*, p. 230, both main text and note 7 on that page.

18. *Foundations of Mind*, p. 361, note 15.

19. *Foundations of Mind*, p. 433.

20. *Foundations of Mind*, p. 360.

21. Saul A. Kripke, *Naming and Necessity* (Harvard University Press, 1980), pp. 113-114.

22. *Foundations of Mind*, p. 360.