



Joseph Margolis

Constructing a Person: A Clue to the New Unity of the Arts and Sciences

I

It is certainly true that no one has demonstrated the sheer falsity of claiming that whatever we count among the most distinctive things of the human world (persons, surely) or that whatever are rightly included among the most salient anthropocentric properties ascribed such things (a capacity for speech and self-reference, for productive and self-transformative agency, and for avowing beliefs, intentions, feelings and the like) are reducible in physicalist terms. Nevertheless, the prospects of a “human science” (a science of the human) confined in reductive terms are very slim indeed – effectively nil. So that *if* the admission of the realist standing of the human world obliged us to weigh seriously the compatibility between a causal theory of human agency and the usual causal canons favored in the physical sciences, we might be forced to concede that deciding what a true science entails would itself be hostage to the contingent difficulties of completing any reductive undertaking here.

The very idea of agency, I suggest, requires an “internalist” rather than an “externalist” causal model. Think for instance of Wittgenstein’s example of “my raising my arm”, which, speaking unguardedly, we say “causes” or “brings about” or (as I prefer to say) “utters” the action in question – which entails but does not cause “my arm’s rising.”¹ We cannot say that an agent’s *raising* his arm causes (in the externalist way) his arm *to rise*, because of course the bodily movement, the arm’s rising, is nothing other than the material event by which the enabling action is itself inseparably realized. The “uttering” of the action and the action “uttered” are never more than internally distinguished within a successful action: they are never more than notionally distinct, never jointly separable in the way the externalist model of causality requires; and the entailed movements of the arm that would ordinarily be identified and explained in the externalist way are no more than the logically dependent subfunctionally factored functioning “parts” of the molar action in question, hence *not* themselves actions in any pertinent sense at all. On the reductionist model, the would-be action should ultimately be no more than a selected set of movements of the kind just conceded, which would effectively retire agency in favor of some Humean-like externalist causal linkage (without reference to agents or persons); and on the agency model, what might otherwise have redeemed the reductionist thesis would now be incorporated, subsumed without distortion, as no more than externally related bodily movements answering to the factored subfunctions of the original action’s molar process, without reference to which *their* causal relevance would remain unspecified.

In this sense, the internalist and externalist models are reconcilable, but only on terms favorable to the agency model; on the externalist model, agency would be abandoned altogether or reduced to a *façon de parler*. It was, in fact, Arthur Danto’s very clever

¹ Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1953), Pt. I, §621.

conjecture (within the first decade of the publication of Wittgenstein's *Investigations*) that suggested that the linkage between actions and bodily movements (in Wittgenstein's example) held the key to how we might understand the conceptual linkage between paintings as artworks and mere canvases covered with paint, which obliquely afforded the perfect clue as to how to construe the conceptual relationship between human culture and physical nature – *a fortiori*, between the human and the natural sciences.² The fact that Danto took a fashionably wrong turn here is no more than a minor complication: the larger lesson rests with the structural analogies disclosed. As far as I know, Danto never explains the difference between nature and culture.

You begin to realize that Wittgenstein's extraordinary question renders the entire fabric of scientific explanation instantly vulnerable in a way it had never seemed before: we are no longer certain what causality means in the physical world or whether it applies in its usual way to human agency; we find ourselves unclear about how to mark the difference between the "natural" and the "human" worlds; and we begin to wonder about what distinguishes science from non-science and what to understand by the very idea of causal explanation. Certainly, there is no prospect of arriving *a priori* at a uniquely convincing picture of "the method of science": Hume and Kant – who, between them, gave the strongest possible reasons for impoverishing our conception of the self and whose influence in this respect may have had a considerable role to play in encouraging, in the 20th century, the return of an analytic bias against enriching that conception – might then be deemed to have been seriously mistaken in their grandest ventures. The methodology of the sciences would then be open to dispute in a very deep way: the idea of the unity of the sciences might remain as insistent as ever but now no longer on the basis of any fashionable model that favored reductionism or an externalist causal model or the primacy of an extensionalist idiom of description and causal explanation or indeed the irreplaceability of the covering law model of explanation.

Danto, I might add, took the wrong turn, not in Hume's and Kant's way by impoverishing our conception of the functional identity of persons but by compartmentalizing (I'm afraid) the analysis of human agency and the proposed reduction of actions to bodily movements; *a fortiori*, the reduction of paintings to painted canvases and speech to uttered sound, as far as numerical identity is concerned. If you allow these analogies to point to the right analysis of the conceptual linkage between culture and nature (which I judge Wittgenstein to have been alluding to) then the would-be reduction of historical deeds to mere bodily movements (by the rhetoric of external redescription applied

² See Arthur C. Danto, "The Artworld," *Journal of Philosophy*, IXI (1964); and *Transfiguration of the Commonplace* (Cambridge: Harvard University Press, 1981). For a sense of Chisholm's extremely intricate efforts at identifying a person's or human agent's "causal contribution" to making something happen, see Roderick M. Chisholm, "On the Logic of Intentional Action," in Robert Binkley *et al.* (eds.) *Agent, Action and Reason* (Toronto: University of Toronto Press, 1971), together with comments by Bruce Aune and Chisholm's reply. Chisholm may well have had Wittgenstein's question in mind at some level of reflection. The essay is clearly a "work in progress." An early incarnation of his view appears in "Freedom and Action," in Keith Lehrer (ed.), *Freedom and Determinism* (New York: Random House, 1966).

I don't subscribe to Chisholm's view but mention him as one of the principal proponents of "agent causation." My own thesis is that agency is a *sui generis* causal model applied to human persons but that persons are not the causes of *their* own actions (or utterances): their own actions are normally not the causes of their further actions; but their actions are the causes of both culturally significant (or culturally "penetrated") events as well as of mere physical effects, and externalist physical causes that are factorial "parts" of *their* actions are also the causes of further physical effects. Normally, we provide reasons (not causes) for supposing that a person has in fact "uttered" a causally potent action. This is an interpretive, not a causal, matter. Accordingly, the explanation of history, artistic production, and practical life is at once both causal and interpretive, which suggests the need to acknowledge the pertinence of an interpretive science.

to bodily movements that we wish to treat as actions) would ineluctably put at risk the coherence of any theory of persons or of the human sciences. There's the threatened *reductio*.³

We cannot manage by merely piecemeal strategies the reduction of action to bodily movement or of paintings to canvases covered with paint or speech to sound and hope to keep our account of the robust existence of persons or selves (ourselves of course) free of reductive risk. All of these bits of analysis must cohere together. Similarly, we cannot insist on the unity of science ranging over the human sciences in the same way the doctrine is said to apply to the natural sciences, as by the positivists, if the theory requires (as it obviously does) an externalist model of causality that could not be applied to human agency unless the agency of human persons were itself reducible – but not otherwise. These interlocking conceptual linkages are much too complex to be treated lightly. Wittgenstein's question cannot be easily answered.

In fact, the realist standing of human persons is well-nigh irresistible, even where it is combated by skillful strategies. Because reductionism in its strictest sense does not actually require the elimination of persons or even the defeat of every form of dualism; because no thoroughgoing reductionism has ever achieved a degree of mastery sufficient to tempt us in the direction of eliminativism; because, for technical as well as practical reasons, the sheer collection of data, our unquestioned reliance on the resources of experience, the proposal and testing of explanatory hypotheses, deliberate commitments to tasks and purposes of consequence make no sense at all without the presence of human persons. This is the larger meaning of P. F. Strawson's well-known objection to the "no-ownership" theory of perception and thought, although Strawson himself is remarkably lax in his account of persons.⁴ It's also the larger meaning of opposing the usual forms of the so-called "supervenience" theory of the mental – Jaegwon Kim's for instance. You have only to think of contrasting Hegel with Kant and Thomas Reid with Hume to appreciate the implied contest.

Strawson's argument and the argument against Kim draw on surprisingly similar strategies, though they are applied in opposite directions and may be rendered well-nigh vacuous. In effect, Strawson's argument, which captures a very strong intuition of what may be called philosophical grammar, holds that feelings, perceptions, thought, intentions and the like must be "adequated" to some existent "subject," agent, organism, or self capable of "possessing" or manifesting "mental states"; such states and occurrences, as of sleep and memory (as we now understand matters), cannot (Strawson argues) be merely contingently or "externally" predicable of the subjects that possess them, though Strawson mistakenly believed that what it was to be a "subject" hardly needed to be elaborated by much to make the point.

Kim's argument, which treats the mental as supervening on the physical in accord with strict causal laws and which supports in that way (and is supported by) the thesis of the "causal closure of the physical" nevertheless fails (for question begging reasons) to come to terms with another intuition of philosophical grammar, namely, that (as in treating a chess move as causally efficacious, in accord with the sense of Wittgenstein's example) a feeling or a thought must not only be possessed by a subject, as an action must be embodied in a physical movement, but the "internal" causal linkages involved can be (dependently) specified *on and only on* the logically prior admission of the would-be supervenient state or occurrence in question – which, of course, implicates the human subject. Thus the material

³ This is the effective upshot of Danto's *Transfiguration of the Commonplace*, Ch. 1.

⁴ See P. F. Strawson, *Individuals: A Descriptive Metaphysics* (London: Methuen, 1959), pp. 95-103; there is an instructive note on Wittgenstein at p. 95n1.

conditions pertinent to the causal efficacy of a chess move are meaningfully specified only as factorially internal “parts” of the would-be supervenient action itself: a chess move may be locally performed in endlessly many ways, though, identified thus, chess moves remain open to informal causal generalizations but hardly to universal or necessary causal laws. Hence, Kim’s analyses put the cart before the horse.

II

Let me put this more forcefully since the problem that Kim considers is ubiquitous with regard to cultural things and to the cultural “penetration” of mind and action (as through language and what language conveys in the way of theory and interpretation and the like): the problem cannot be confined to the biology of the mind. I don’t believe Kim, who may well be the most skillful and unyielding reductionist of the current Anglo-American analytic movement, ever offers a reductionist account of languaged thought or speech; and yet, of course, without success there, reductionism would be dead in the water.

Apart from that, Kim is committed among his most recent discussions to the following thesis, which he calls “*conditional physical reductionism*, the thesis that if mental properties are to be causally efficacious, they must be physically reducible.” Now, this doctrine is meant to provide an answer to the problems of “mental causation and consciousness,” which, effectively, *construes* these questions in a peculiarly restricted way:

Each poses a fundamental challenge [Kim concedes] to the physicalist worldview. How can the mind exercise its causal powers in a causally closed physical world? What is there, and how can there be, such a thing as the mind, or consciousness, in a physical world?⁵

It’s important to understand that the distinction of the cultural world and of the human sciences and studies that address that world *requires* the defeat (or stalemate) of Kim’s version of reductionism (which affects the fortunes of a very large swath of familiar forms of the doctrine); but then we must also see why supervenientism fails on its own terms.

The answer is straightforward but needs some care. It’s worth remarking that Kim’s solution is not troubled by any hint of dualism; on the contrary, Kim regards his thesis as made stronger by its compatibility with dualism: because of course dualism would have no adverse bearing on the causal question *if* Kim’s “conditional physical reductionism” were true.

That would count as an interesting gain – compatible for instance with epiphenomenalism and the classic forms of emergentism. It helps to explain why Kim’s usual treatment of the “mental” (wherever the mental plays a causal role) deliberately yields in the dualist direction – though not of course in any way meant to agree with Descartes’s causal doctrine. Tangential questions, you realize, threaten to overwhelm us here: we must keep our discussion of the mental as close as possible to the cultural “penetration” and transformation of our biological gifts; we must hold fast to the agency of persons; and we must keep before us the differences between the human and the natural sciences. These are our principal touchstones.

The trouble with Kim’s supervenientism is that it completely neglects a natural option (a stronger option in my opinion) that arises in connection with the distinct kinds of

⁵ Jaegwon Kim, *Physicalism, or Something Near Enough* (Princeton: Princeton University Press, 2005), pp. 5, 13.

evolution separately implicated in consciousness (or mind) and in the cultural world. For one thing, if dualism is a conceptual scandal both metaphysically and causally, then it's more reasonable to treat the evolution of mind and culture *non*-dualistically if we can – and of course we can. And for another, it's reasonable to think of the evolution of mind as thoroughly biological, but not the evolution of culture – hence not the cultural transformation or *Bildung* of the mind.⁶ For even if the “mental” and the “cultural” possess *physical* features (as perception, thought, and speech do), Kim's entire argument would still risk irrelevance. Indeed, it's not only possible – it's true – (i) that there are events that we characterize as mental or culturally significant that possess physical features of their own and produce effects in the physical world that cannot be convincingly replaced by mere physical sequences: a verbal insult, say, that produces anger and the reddening of the face; and (ii) that what is “emergent” here in some pertinently evolutionary sense is *not* “supervenient” according to Kim's formula, because of course supervenience is explicitly dualistic whereas cultural emergence is not and because the material realization of the culturally emergent is logically inseparable from what is actually emergent. How could it be otherwise?

If you grant all this, you grasp the sense in which Kim's argument relies on a deep equivocation that he seems never to have addressed or to have thought needed to be addressed. He's missed the most important possibility! Here's the evidence:

Mental properties [Kim says] supervene on physical properties, in that necessarily, for any mental property M, if anything has M at time t, there exists a physical base (or subvenient) property P such that it has P at t, and necessarily anything that has P at a time has M at that time.⁷

But this cannot be true or even relevant if there are no psychophysical laws or reductionist laws by which to validate the last clause of Kim's formulation. But there are no laws linking the cultural and the physical, or mental powers that are culturally penetrated – because in introducing cultural events we already make provision for those *subfunctional* physical events by which the cultural is duly realized. In fact, quite independently, there is no known argument to show that there *are* any necessary or exceptionless causal laws at all, or that the commitment to exceptionless laws cannot be abandoned without loss.⁸

Admitting the larger argument against reductionism obliges its advocates and allies to muster a better campaign than they thus far have. Strawson himself, conceivably because he failed to distinguish (in *Individuals*) between his own (intended) theory of persons and

⁶ See, for a firm admission that the propagation of culture cannot be accounted for in terms of Darwinian evolution, Richard Dawkins, *The Selfish Gene*, rev. ed. (Oxford: Oxford University Press, 1989); see, also Mario Bunge, “Emergence and the Mind,” *Neuroscience*, XI (1977), for a sketch of an emergentism that fails to consider viable forms of emergence that, like the evolution of the cultural world, are relatively independent of the organization of physical and biological systems on which they nevertheless build.

⁷ Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge: MIT Press, 2000), p. 9. It's fairly clear that Kim takes “supervenience” and “emergence” to be pretty nearly equivalent. There are such uses, but it is not in the least clear why Kim does not consider emergentisms that eschew dualism altogether and the unyielding doctrine that there must be exceptionless covering laws for all causal sequences. See, for instance, Jaegwon Kim, *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press, 1993), pp. 134-135. See, also, Lloyd Morgan, *Emergent Evolution* (London: Williams and Newgate, 1923), cited by Kim in preparing his argument here. The *Physicalism* and *Mind* texts go no further. Kim's modest finding in all this comes to this: “it seems clear to me that preserving the mental as part of the physical world is far better than epiphenomenalism or outright eliminativism,” *Physicalism, or Something Near Enough*, p. 120. Yes, of course, but those are hardly the important options.

⁸ See for instance the strong argument offered in Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon, 1983).

an unsatisfactory dualism, or (say) a self-styled hermeneut like Charles Taylor, who never quite realizes that the choice between reductionism and the hermeneutic vision cannot for the strictest of reasons assume a disjunctive form, cannot be an effective guide.⁹ The reason is plain: a consistent reductionism can perfectly well admit the entire human world and (as Kim argues) at least some standard forms of dualism, and still coherently seek to “reduce” the description and explanation of its features in materialist terms. That is in fact reductionism’s charge. Taylor neglects this elementary truth. In fact, given the history of what counts as “matter” in the physical sciences, it would not be beyond reason to recommend we dub the mental and the cultural as “material” phenomena themselves – obviating thereby the reductionist’s supposed advantage at a stroke.

But if you grasp all this, you cannot fail to see that the same question confronts us when we worry whether reductionism affords the right relationship between the natural and the human sciences or between the arts and the sciences or between theory and practice. The reason is simply that, *prima facie*, human persons *are* the ineliminable agents of all the arts and sciences – the “middle term,” so to say, of any argument recommending a redefinition (modest or radical) of what, philosophically, to understand an art or science to be. And there, though it may seem otherwise, the theory of the human self, the paradigmatic agent or subject of action and utterance – or of thought, perception and feeling, or of purpose, intention and commitment, or of responsibility, interpretation and appreciation, or of technology and creativity – is a theory about one and the same being. I don’t mean to disallow eliminativism in principle. But certainly, it would be absurd to ignore the fact that, in denying that there *are* persons, that is, existing persons – a thesis Wilfrid Sellars manfully tried to show us a way of eliminating¹⁰ though it’s very likely Sellars’s effort was meant to be no more than a thought-experiment, possibly a joke, since he willingly restored persons and what he called their “intentions” to their ordinary niches, by trivial rhetorical “addition” – *we* would still be bound to deny our own existence!

III

There’s a deep puzzle there. Strawson’s objection to the no-ownership theory may be reasonably reinterpreted as a grammatical rather than an explicitly metaphysical thesis, meaning by that that what we informally call “the mind” is intended to collect our sense of the functional coherence, even unity, of a set of distinctive attributes instantiated in the lives of human beings: “mind” is the nominalization of that sort of functional unity, whatever may prove to be a perspicuous theory of the “I” or “self” or “soul” of the human world or its animal or (possibly) its machine surrogates among robotic selves. Derek Parfit, for instance, in what may be the first of his penetrating efforts to define what is minimally necessary in theorizing about the “subject” of thought, experience, memory, and action, failed to concede the full force of Strawson’s complaint – which, rightly applied, counts as a severe objection to both Hume and Kant (though for very different reasons).¹¹ Parfit disjoined attribute and referent grammatically (or logically) when he rightly objected to excessively grand theories of some substantive “I” ranging over the whole or large phases

⁹ On Strawson’s difficulties, see Bernard Williams, *Problems of the Self* (Cambridge: Cambridge University Press, 1973). See, also, Charles Taylor, *Philosophical Papers*, 2 vols. (Cambridge: Cambridge University Press, 1985).

¹⁰ See Wilfrid Sellars, “Philosophy and the Scientific Image of Man” and “The Language of Theories,” *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963).

¹¹ See Derek Parfit, “Personal Identity,” *Philosophical Review*, LXXX (1971); and *Reasons and Persons* (Oxford: Clarendon, 1984).

of human life. Nevertheless, at least two very ardent eliminativists, Daniel Dennett and Paul Churchland, were, I think, literally convinced by Sellars's arguments.¹² Not, I may say, Sellars himself, as far as the evidence goes.

Also, in admitting the "existence" of persons we are not yet obliged (let it be noted) to hold that persons are or are not substances or entities of any canonically familiar sort (the Aristotelian sort, for example). Perhaps "person" need signify no more than the notional site or seat of certain culturally emergent competences that cannot be described or explained in terms of the merely natural biological powers of the animal species *Homo sapiens*. "Person" and "organism" may well be conceptually incommensurable distinctions though *not* for that reason incompatible categories affecting the would-be unity of the sciences. In any case we cannot rightly give an account of the "unity" of the sciences, or the unity of the arts and sciences or of theoretical and practical concerns, without a robust theory of the agency of persons, unless we turn (unwisely, as the evidence suggests) in the direction of reductionism.

The logic is clear but the metaphysics is disputed. It's worth pondering, therefore, why the admission of intelligent animals – dogs, elephants, chimpanzees, dolphins – does not oblige us to exceed the functional coherence and unity of such creatures' lives in the direction of the conceptually swollen theories usually reserved for humans. The obvious answer concerns the *sui generis* emergence of the cultural and the cultural penetration (and artifactual transformation) of our animal powers and the absence among the most intelligent animals of more than an incipient form of proto-cultural learning too weak to give evidence of any artifactual "I." It's the artifactual "I," which the mastery of language entails and makes possible, that compensates more than adequately for the impoverished theories of Hume and Kant. The point is that confirmation of the presence of the "I" *is* empirical, though not phenomenal in the empiricist sense or, *a fortiori*, rationally necessary in the transcendental sense. We can actually witness the growth of the "I" among our children.

Persons or selves, I would say, are, paradigmatically, the artifactual sites of our culturally acquired powers, exerted in whatever transformative ways they are through whatever is biologically given by our membership in *Homo sapiens*. It may then be ultimately more important to enrich our sense of the functional powers of persons than to speculate on any "substantial" difference between mind and matter. Kant's surprisingly misguided elaboration of what he calls the "concept or judgment" of the "I think" (his transcendental revision of Descartes's *Cogito: Ich denke*) – somehow added to (or made to "accompany") the putatively *completed* system of his transcendental categories – surely counts as an embarrassing lesson. But if that is true, then so, too, is Dennett's impoverishment of the human "mind," where what may be needed are promising thought-experiments capable of eclipsing eliminativist options¹³.

It comes as a surprise to discover how entrenched an animus there is against persons or selves in modern Eurocentric philosophy: either against their very existence (eliminativism) or against their possessing a perceptually discernible nature matching their apparent form of life – as, among 18th century theories, following the decline of rationalism, in the views of figures like Hume and Kant: after failing in the relevant regard, empiricism and transcendentalism have led the analytic temperament back to reductionism and dualism –

¹² See Daniel C. Dennett, *Content and Consciousness* (London: Routledge & Kegan Paul, 1969) and *Consciousness Explained* (Boston: Little Brown, 1991); and Paul M. Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (Cambridge: MIT Press, 1990).

¹³ See Dennett, *Consciousness Explained*, Chs. 9, 13, 14. Compare Bernard J. Baars, *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press, 1988).

our prevailing contemporary preferences. Hume cannot find any empiricist datum to count as the “I” of any of “us” – rightly so; and Kant is all but defeated by having to attach to his perfectly closed system of categories the external, oddly fitted, almost entirely undefended and unexplained function of the “I think,” which he renders as a surd that has no other purpose but to “accompany” or “introduce” (as he puts it) the categories themselves applied to sensuous intuitions – so as not to produce unwanted paralogsms.

We must remember not to impoverish our account of the human mind in our zeal to favor one or another theory of the various sciences and arts, or practical and theoretical concerns. There you have the decisive clue to the strategic importance of our conception of persons or selves in seeking a rapprochement between (say) the analysis of painting and literature and the analysis of physical processes and the existence of human societies. On its face, it seems preposterous to suppose that the description, interpretation, explanation, and appreciation of what obtains in the arts and sciences would never require the robust role of a reflective “I,” however subject it may be to historied conviction and evolving experience and interest. The 18th century was doubly victimized by the effective absence of the conceptual resources of modern evolutionary theory and the historied *sui generis* emergence of the cultural world from the biological: without these, it’s quite impossible to construe the human “I” as an artifactual achievement; and without that extraordinary invention, the entire puzzle of the human sciences would make no sense at all.

By the close of the 18th century, Hume and Kant take command of the two principal forms of subjectivism – one psychologistic, the other not – which in a curious way are inseparable from one another and plainly dominate a very large part of the subsequent history of philosophy down to our own day. Their theories, however, have impoverished our picture of the human subject, and have, as a consequence, provoked a profound reaction among the post-Kantian Idealists and their immense progeny ranging among the existentialists, the Marxists, the pragmatists, the champions of *Lebensphilosophie*, the hermeneuts, the Nietzscheans, the phenomenologists, the Freudians, the Heideggerians, the advocates of *Weltanschauungsphilosophie*, the Frankfurt Critical school, the Wittgensteinians, and others, who hurry to re-enchant the world by restoring an enhanced account of what it is to be a human person.

Here, I find the most important, most inventive sources of the philosophical recuperation of the human to be the following two 19th-century contributions: namely, Hegel’s notion of historicity and Darwin’s theory of evolution. Without pausing to explain for the moment the terms of art I favor here, let me say that Hegel provides the most important new conception of what may be termed “internal *Bildung*,” a notion (still borrowing from the German to explain the Greek) akin to Aristotle’s themes of *sittlich* education, as in his *Ethics, Politics, Poetics, and Rhetoric*, except for the fact that *Bildung* must be construed as a specifically enculturating form of instruction under the condition of historicity – distinctions Aristotle was all but completely unaware of, which surface incipiently in the 18th century in Vico and Herder and find their first great conceptual articulation in Hegel’s remarkable achievement.¹⁴

¹⁴ You will find a somewhat inchoate grasp of the *sui generis* world of human history and culture – in philosophical terms – in John McDowell’s *Mind and World* (Cambridge: Harvard University Press, 1994, 1996). See, also, *Reading McDowell: On Mind and World*, ed. Nicholas H. Smith (London: Routledge, 2002). McDowell tries to use Aristotle’s and Kant’s conceptual resources in order to capture something close to Hegel’s and Gadamer’s notions of *Bildung*, but the effort is abortive. That the effort was made as late as the end of the 20th century by one of the most promising younger figures of Anglo-American analytic philosophy is something of a surprise. I think one cannot make sense of it without conceding that the influence of Hume’s and Kant’s conceptual economies (with regard to the characterization of the self or human subject or person) are as strong as they ever

Hegel's theory is an expression of high philosophy, but Darwin's is not. Darwin provides the essential empirical grounds for the elaboration of what (by a term of art) I shall call "external *Bildung*," meaning by that the gradual evolution of prehuman and proto-human modes of intelligence and communication close to achieving the rudiments of true language and the forms of self-reference and self-identity and other *sui generis* skills that (we take to) *constitute* the first appearance of those all-important cultural artifacts – those "second-natured" hybrid creatures – we call persons. Of course, affirming all this counts as issuing a very large promissory note. But I must offer a more pointed sense of the novel "entity" I'm enlisting before we allow ourselves to become completely occupied with its defense.

I take persons to be a certain kind of cultural construct, enjoying realist standing as such, which, if true, could not have been rightly defined philosophically until about the turn into the 19th century or after Darwin. This literally means that, for somewhat more than the first two millennia of the whole of Western philosophy (which is itself nearly the whole of Western philosophy), it was literally impossible to formulate a reasonably correct "metaphysics" or "philosophical anthropology" of the human. I find that a stunning admission, closely related (in my mind) to the explanation of Plato's futile attraction to the Forms when, defining the virtues in the elenctic dialogues, Socrates clearly lacks any conceptual resources that might have made it possible to avoid admitting the Forms – in proposing (say) the radical idea of the cultural construction or constitution of the virtues themselves, a thesis that very naturally matches the idea of the cultural construction of persons. Think of the paradox of attempting to explain the Socratic virtues in a Darwinian world in which the human species had not yet evolved! To grasp the lesson is to grasp the insuperable impoverishment of rationalist, empiricist, transcendentalist, and Idealist options: we cannot advance beyond one or another constructivist constraint.

Plato clearly has the use of the concept of *paideia* (internal *Bildung*) but not external *Bildung* in anything like the sense that would admit the primordial appearance of the human. I find it suggestive to think of persons as "natural artifacts," meaning by that that, in their merely biological niche, they are incompletely formed for their characteristic cultural role; hence also, that their biological gifts prepare them for their "second-natured" enculturation: their mastery of a home language, for example.¹⁵ The human sciences are focused on the *sui generis* powers of a hybrid being that is "artifactual by nature" (that is, by becoming "second-natured"). If then you take the liberty of characterizing mind and culture as "material" – meaning to equate the natural and the material uncontroversially – you realize that you've stalemated reductionism, dualism, and eliminativism at a stroke without yet declaring what the distinction of the human world actually is.

Let me put the point a little more quarrelsomely. Aristotle, I suggest, never needed to invoke what I'm calling "external *Bildung*" because, whatever his temptation, he never seriously exceeded a *sittlich* reading of the normative (if I may put the matter thus); whereas, already in the *Republic*, Plato pursues the supposed discovery of the ultimate Forms by which all questions of right conduct and right belief are said to be governed. Hence, Aristotle constructs a reasonable picture of the good life and the good *polis* largely in terms of his comfortable attraction to Athenian life; similarly, a reasonable picture of the best of Greek tragedy, in accord with his preference for Sophocles. Whereas Plato makes it

were in the 18th century. For a glimpse of Hegel's sense of "Bildung," see his *Phenomenology of Spirit*, trans. A. V. Miller (Oxford: Oxford University Press, 1977), Introduction.

¹⁵ I take the term from Marjorie Grene, "People and Other Animals," *The Understanding of Nature: Essays in the Philosophy of Biology* (Dordrecht: O. Reidel, 1974).

clear that, on the strength of Socrates's project (in the *Republic*), Homer and the most admired poets will have to be exiled from the ideal state. *If* the Forms must and can be contested, then, as history makes clear, we cannot fail to address the question of "external *Bildung*," since the choice of supposedly real norms of goodness and truth would still confront us if we conceded that truth and goodness must have an artifactual provenance. I view that as a radical option to which we are inexorably driven.

IV

Once you glimpse the force of this last challenge, you begin to see the extraordinary paradox produced by Hume and Kant in impoverishing (in their very different ways) the notion of the "I," the notion of the subject and agent of everything distinctly human. As I say, the self becomes a surd in Hume's official philosophy: Hume retreats very cannily (when he wishes) to the idiom of ordinary humanity when, very sensibly, he overtakes the scandalous threat of the "no-ownership" thesis; for his part, Kant enriches the transcendental subject's judgments regarding the standing of Euclid's geometry, moral obligation, and the disinterested pleasure of natural beauty; but he never quite acknowledges how much more in the way of conceptual resources he actually needs in order to account for instance for at least the history of science (if not also the history of morality and an engagement with the fine arts) and what in particular he needs to acknowledge transcendently about the "I"'s agency apart from the categories and pure intuitions the "I" applies to sensuous intuition. He simply does not assign sufficiently detailed powers to his transcendental ego to accomplish its usual chores even with respect to the principal kinds of judgments he examines – or to explain how those powers arise.

Kant's third *Critique* provides the most notorious evidence in support of the charge, particularly if you read its argument as an emendation of the first *Critique*. There's next to nothing in the first part of the third *Critique* – bearing on the resources of the mind – to account in transcendental terms for any familiar creative or critical practice regarding the fine arts – or "language" or "history" or "culture" in the large. Kant is driven, for example in his gymnastic speculations about new ways to trick the faculty of imagination into applying the concepts of the understanding to our interest in artworks (or our appreciation of beauty in nature of course), in a way that would not violate his well-known taboo against treating aesthetic judgments as cognitive in any sense. This is a hothouse quarrel, I concede, but it yields a compelling lesson nevertheless. I'm interested here not so much in the local puzzles of moral and aesthetic theory as in the consequences (for any compliant philosophy) of Kant's having impoverished our conception of the "I" – *and* his having failed for that reason to realize how doing *that* renders every part of his own philosophy conditionally suspect and arbitrary. Hard as it is to believe, Kant seems not to have noticed that the entire structure of the first *Critique*, preeminently the strategic function of the closed system of the categories and pure intuitions of the understanding, depend on the legitimate cognitive powers of the "I" rather than on the seeming sufficiency of our guesses at the completeness of his proposed run of fundamental categories. In any case, you cannot have the one without the other.

We're speaking here of the most influential philosophical mind of the last two and a half centuries. Kant simply abstracts all that he can say about his transcendental subject (the "I think") *from* whatever, *independently*, can be derived from his account of the various kinds of judgment he allows; but these are themselves drawn up only dialectically – that is, *from* a quarrelsome literature, *not* from any examination of the salient

competences manifest in actual practices in any of the arts or sciences.¹⁶ In fact, the “free play of the imagination” featured in the third *Critique* has led figures like Wilhelm Dilthey and Ernst Cassirer to weigh the possibility that Kant may have signaled the need for a looser account of the “system” of the categories than he offers in the first *Critique* – in order, precisely, to account for the historical nature of human being itself (Dilthey) or the emergence of novel “symbolic forms” (Cassirer) that may not be able to be accounted for on the basis of Kant’s original system.

I am concerned here to demonstrate how impossible it is to justify any plausible account of the admitted work of the sciences and arts without a ramified theory of the nature of the human self. I’ve dwelt on Kant’s and Hume’s theories of the self in order to remind us of just how impoverished a picture Western philosophy was obliged to draw on by the end of the 18th century (at the very dawning of “modern” modern philosophy, ushered in by Kant’s great revolution) and how through the work of analytic philosophy in the 20th century we have somehow returned once again to the impoverishment of the concept, signaled in different ways (as I’ve suggested) by strategically placed figures like Wilfrid Sellars and Charles Taylor.

Let me clinch the point, therefore, by merely citing what Kant offers in abruptly introducing the idea of a cognizing function, the “I think,” which he calls a “concept” or “judgment” but which has no other function to perform except what is confined as follows:

The I think [Kant affirms] must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me

It is only because I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself, i.e., the analytical unity of apperception is only possible under the presupposition of some synthetic one.¹⁷

There’s the *deus ex machina* of Kant’s first *Critique*: all of the analyses of the sciences and arts, of theoretical and practical matters, of judgment and sensibility, are formulated (by Kant) without sustained or direct attention to whatever is problematic about experience or practice or the influence of history or prejudice or perspective or the *Bildung* of a human being. This, of course, was Herder’s prescient worry and the focus of Hegel’s profound correction.

But apart from these instructive details, the most forceful demonstration of what must be recovered belongs decisively to Ernst Cassirer, though the lesson’s already obscurely adumbrated in Kant himself and among the post-Kantian Idealists and their progeny. Cassirer’s “philosophy of symbolic forms” is at once a Hegelianized eclipse of the adequacy of Kant’s transcendentalism and an attempt at recouping the Hegelianized possibilities of a Kantian-like reformulation of the “I think” that effectively admits the *a posteriori* standing of the resultant transcendental definition of the “categories” of “all” our *Wissenschaften* (sciences, studies, arts, technologies) constructed under the condition of history and historicity. Cassirer “commits” himself to Kant’s transcendental outlook but does not actually “demonstrate” the necessity of the transcendental categories. In effect, he

¹⁶ I pursue the theme – with regard to the philosophy of art – in *Aesthetics: An Unforgiving Introduction* (Belmont: Wadsworth, forthcoming 2009).

¹⁷ Kant, *Critique of Pure Reason*, §16.

converts Kant's "critique of *reason* [into] a critique of *culture*."¹⁸ The truth is, Cassirer *supersedes* the Kantian categories by introducing an openended evolving set of "symbolic forms": most tellingly perhaps in his account of the late phases of modern physics. He's aware of this (and admits the fact obliquely), but avoids a direct confrontation with the Marburg Kantians. There's the upshot of the Kantian heritage for our time.

Which is to say: the resources of the productive or creative agency of the "I" cannot be construed abstractly (in the manner of Kant's first *Critique*) but must follow the actual exemplars of how history and experience are concretely exploited. In effect, the entire fund of cultural history informs our picture of the inventive powers of the "I." That is precisely what Hegel attempts to gather in his critique of Kant: what cannot possibly be defined or transcendently confirmed according to Kant's conception. It would not be unreasonable therefore to read Kuhn's *Structure of Scientific Revolutions*, Nietzsche's *Genealogy of Morals*, Hegel's *Aesthetics*, and Cassirer's *Philosophy of Symbolic Forms* as pointed reflections on the functional powers of the "I think" in the segregated spaces of science, moral reflection, the aesthetics of fine art, history, mythic imagination, religion, technology, semiotics, interpretation, and whatever other sectors of human interest Kant fails to feature; and to see in their kind of contribution the need to oblige Kant's transcendental inquiries to address the very problem he bruits in acknowledging the nearly completely neglected role of the "I." The analysis of the "I" and the analysis of its powers are inseparable.

Seen this way, there is and can be no principled disjunction between concepts and categories: the *a priori* is no more than *a posteriori* while remaining a second-order conjecture; there cannot be any universally adequate closed system of the categories of description and explanation; general predicates make sense only in the context of their provisional exemplars, which must themselves be continually replaced with evolving experience; every vestige of strict universality and substantive necessity must and can be retired; inquiry (of every kind) must be inherently openended, subject to potentially radical revision however regulated by our evolving notions of the relative adequacy of our experience and theories; and the unity of all such efforts, whether among the sciences or between the sciences and the arts, cannot but depend on our theories of how persons are culturally constituted and what we take their evolving "natures" to be.

V

In this sense Cassirer's penchant for the idiom of universality and necessity (a Kantian disorder, let us say) must be set aside in the same way Hegel's extravagances are, by redefining the continuum of the finite and the infinite in terms of the moving but insuperably inaccessible "limit" of human inquiry, which bids fair to facilitate (in a new way) the unity of the arts and sciences. The prototype of this mode of reconciliation, incipiently perceived among the post-Kantian Idealists, was perhaps most perspicuously remarked, in terms of rational hope more than of accessible knowledge, somewhat later than the middle of the 19th century, somewhat after the publication of Darwin's *Origin of Species*, by that oddly transplanted German Idealist the American pragmatist Charles Sanders Peirce, whose theories Cassirer was aware of and in certain important respects unexpectedly shared.¹⁹ Cassirer simply restricts the more strenuous Kantian idiom to the

¹⁸ Carl Hamburg, "Cassirer's Conception of Philosophy," *The Philosophy of Ernst Cassirer*, ed. Paul Arthur Schilpp (La Salle: Open Court, 1949), pp. 77, 86.

¹⁹ See my "Rethinking Peirce's Fallibilism" *Transactions of the Charles S. Peirce Society*, XLIII (2007).

compartmentalized space of his “symbolic forms”; but he does not treat the “forms” as demonstrably necessary categorial constellations – or as “complete.”

In any case, Cassirer locates a particularly strategic – in fact a decisive – piece of evidence regarding the physical sciences that confirms the inherent provisionality of Kant’s entire table of transcendental categories: which, once admitted, marks the historicity of explanatory theories among the physical sciences, the *a posteriori* standing of transcendental apriority (when separated from transcendentalist presumptions); and, most important, the need for an enriched account of the “nature” of human agents and inquirers, which ranges over the whole of human life (including the arts, religion, and myth) often against the scientific models of science that still prevail.²⁰ All of this bears directly on the prospect of reconceiving what it is to be a science.

Cassirer himself seems to have believed – it’s the nerve of *The Philosophy of Symbolic Forms* – that the model of transcendental necessity and universality in Kant’s first *Critique* is resilient enough to accommodate the “classical” system of Galilean-Newtonian dynamics [and] the principles on which it rested [drawn from “observed phenomena” in such a way that] the basic framework of physical reality seemed to have been established once and for all [as well as that deeper] revolution in method in the field of physical theory [first developed by Heinrich Hertz and Wilhelm Helmholtz, which turned away] from the copy theory of physical knowledge [based on sensory representation] to [favor] a purely symbolic theory.

That is, turned away from confining explanatory theory to the limited data of sensory perception (applied to conceptually instantiated *explananda* and *their matched explanantia*) in order to treat perceptual data in a freer way – as “signs” of their compliance *with* independent, mathematically invented abstract functions, no longer constrained by the *initial* categories governing sensuous intuition in Kant’s sense (but to which, within the resources of some present history, the observable data could then be shown to conform).²¹ The upshot of this single, seemingly small adjustment is philosophically electric. It shows unmistakably that Cassirer eclipsed the supposed “completeness” of Kant’s categories even as he presented himself as an orthodox neo-Kantian: it marks the nerve of the radical revision, within the Kantian movement, of the very idea of a closed system of description and explanation.

If you consider the transformation Cassirer carefully flags here, you begin to see that his own loyal and generous adherence to the Kantian philosophy – or to as much of the Marburg interpretation as he dared concede – may already be too sanguine for the Hegelian innovations he’s also adopted. They already implicate a conception of an active subject – an “I think,” if you like, or, better, an “I” capable of inventing in the freest way explanatory proposals of a hitherto unforeseen “transcendental” sort, well beyond the assured ken of Kant’s “completed” table of transcendental categories (and his “I think”). That is part of the upshot of Cassirer’s conception of “symbolic forms” applied particularly among the

²⁰ See my *The Unraveling of Scientism: American Philosophy at the End of the Twentieth Century* (Ithaca: Cornell University Press, 2003).

²¹ Ernst Cassirer, *The Philosophy of Symbolic Forms*, vol. 3, trans. Ralph Manheim (New Haven: Yale University Press, 1957), p. 20.

²² The best (brief) demonstration of the now-contingent nature of the Kantian *a priori* that I’m familiar with appears in C. I. Lewis’s classic paper, “A Pragmatic Conception of the *A Priori*,” *Journal of Philosophy*, XX (1925). The argument, therefore, takes a pragmatic turn, which is precisely what explains the strong convergence between the views of Peirce and Cassirer.

²³ See, for instance, A. G. Cairns-Smith, *Genetic Takeover and the Mineral Origins of Life* (Cambridge: Cambridge University Press, 1982).

sciences and judiciously extended to the spheres of myth, art, language, history, and the like, regardless of Cassirer's "official" use of Kant's transcendental idiom.²²

My own conjecture has it that what must be added *here* cannot be slimmer than whatever may be required in admitting creativity and agency among the arts, technology, philosophy, and practical life. There's the point of pursuing Kant's failure to fulfil his own transcendental goal. This may be fairly counted as the beginning of "modern" modern philosophy, sketched first in Hegel's classic anticipation in the *Phenomenology* – however unmanageably for his successors. Once you admit that the sciences are artifactual constructs of a kind no longer minimally restricted to the boundary conditions of sensory experience (either descriptively or explanatorily) and open, as a consequence, to a new sort of confirmation under conditions of historicity prepared to abandon all presumptions of necessity and universality, there can be no antecedent limitations on explanatory concepts and categories that could possibly vindicate Kant's original convictions regarding transcendental closure or the limitations Kant imposes on the "I think." Kant's picture is already too thin for Cassirer's needs!

The fact is that the original appearance of organismic life may have been due to contingencies of a fortuitous and probabilistic sort that cannot be shown to depend on causal laws of the classic kind.²³ Universality and necessity may be conceptual extravagances that finite inquiry cannot validate – however useful such idealizations may prove to be. Here, the contingencies of the arts, politics, natural-language discourse, history, and the adequacy of institutional practices fitted to all the contexts of human interest militate against the presumption of ever discovering "the" changeless order of things. It's in this sense that the new unity of the arts and sciences depends on the cognitive complexity of the role we assign the human subject in all our inquiries. Cassirer has put his finger on a decisive *reductio* of the Kantian orthodoxy he does not directly attack.

There are indeed some telltale oddities in Kant's first *Critique* that (against the *Critique's* intention) hasten the eclipse of Kant's version of his own transcendental revolution. If, for instance, you treat Euclid's geometry and Newton's physics as the natural sites of Kant's way of proceeding philosophically, then it's quite impossible to make compelling conjectures about the necessity, universality, fixity, or completeness of Kant's table of basic categories – except conditionally, relative to whatever is generated in the actual history of science: the contingent successes of what Kant calls the faculty of understanding (*Verstand*) must then be seen to introduce no more than a run of empirically contingent concepts for which, on Kant's own say-so, reason (*Vernunft*) is expected to discern those deeper universal categories that necessarily inform whatever concepts belong to any particular interval of historical adequacy. But the idea was wildly off the mark with respect to both non-Euclidean geometry and relativity physics – and, as we now realize, it had to be. It takes but a moment's reflection to grasp the telling fact that historicity and universality are bound to be finally incompatible.

If the theory-laden "givens" of perceptual experience are themselves emergently or historically contingent, then the transcendental universals under which they are rightly subsumed cannot be more than conditional as well. Consider only that we seem capable of endlessly improvising new concepts and categories to match evolving experience – at precisely the same time that experience obliges us to make the effort! How could we possibly claim that any closed system of categories would be sufficient for all time? Nevertheless, it's here, precisely, that Kant reverses the seemingly sensible order of things

²⁴ Kant, *Critique of Pure Reason*, B143-144.

²⁵ See Kant, *Critique of Pure Reason*, §§20-22; compare §16.

– in order to ensure the apriorist standing of his own thesis. He holds (as, in effect, we've seen) that “the categories arise independently from sensibility merely in the understanding,” and that in this sense “the manifold that is given in a sensible intuition necessarily belongs under the original synthetic unity of apperception, since through this alone is the unity of the intuition possible.”²⁴

Kant never shows that this must be true, and indeed, given Cassirer's argument, the claim is defective on its face. For the “I think” ensures the unity of the categories merely by introducing “the logical function of judgments” and confirms (tautologically) that the completeness and closure of the system of our categories are rationally assured, since otherwise no “deduction” of the categories could claim its necessary standing.²⁵ Extraordinary trickery! That is, Kant does not grasp the falsity of his own constraint on the intelligibility of the physical sciences for the same reason he cannot grasp the conceptual limitation of the Newtonian picture of a valid physical science (one, say, that claims not to “feign hypotheses”): he was unable to anticipate the innovations of Hertz and Helmholtz in exactly the same sense in which he could not imagine a non-Euclidean geometry. But once grant that, we cannot fail to see that the categorial requirements of all of Cassirer's symbolic forms must be constructed and empirically testable rather than transcendently confirmed in Kant's apriorist sense.

Cassirer's discovery develops along entirely different lines from Kant's own philosophical strategy. What Cassirer demonstrates in taking account of modern physics' eclipse of the “copy theory” of science (which in Kant's book presupposes the passivity or receptivity of sensibility)²⁶ is nothing less than the knockdown evidence that the mathematized thought of modern physics (the *a priori* construction of mathematical “objects,” say) is indeed explanatorily productive just where it frees itself from a “reliance on sensuous or perceptual guides.” That is precisely its “intellectual triumph.”²⁷

²⁶ Kant, *Critique of Pure Reason*, A19-20-22/B33-34.

²⁷ See Harold R. Smart, “Cassirer's Theory of Mathematical Concepts,” in Schilpp (ed.), *The Philosophy of Ernst Cassirer*, pp. 256-259.

²⁸ See Thomas Nagel, “What is it like to be a bat?” *Philosophical Review*, IV (1974); and David Chalmers, *The Conscious Mind* (New York: Oxford University Press, 1996) and “Facing Up to the Problem of Consciousness” and “Moving Forward on the Problem of Consciousness,” in *Explaining Consciousness – The ‘Hard Problem’*, ed. Jonathan Shear (Cambridge: MIT Press, 1995-1997).

²⁹ See Chalmers, “Facing Up to the Problem of Consciousness,”

p. 10; compare Charles Taylor, “Self-interpreting Animals,” *Philosophical Papers*, vol. 1 (Cambridge: Cambridge University Press, 1985).

³⁰ Chalmers, “Facing Up to the Problem of Consciousness,” p. 11.

³¹ See, for instance, Edmund Husserl, *Ideas: General Introduction to Pure Phenomenology*, trans. W. R. Boyce Gibson (New York: Collier Books, 1962), “Author's Preface to the English Edition.” Hegel's and Peirce's phenomenologies, let it be noted, are completely unrestricted (as against privilege) when compared with Husserl's model.

³² Cassirer and Peirce hold similar but not identical accounts of the triadic nature of representations, signs, symbols, and other semiotica. Cassirer is somewhat more explicit than Peirce about confining semiotic “objectivity” to what obtain *within* the space of particular symbolic forms – which corresponds to his analysis of the relative autonomy of mathematized “objects” introduced in the physics of Hertz and Helmholtz; whereas Peirce never entirely resolved the relationship between a realist and idealist account of science itself. See, here, Hamburg, “Cassirer's Conception of Philosophy,” pp. 82-85; Ernst Cassirer, *The Philosophy of Symbolic Forms*, vol. 1, trans. Ralph Manheim (New Haven: Yale University Press, 1953), “Introduction and Presentaton of the Problem”; Susanne K. Langer, *Philosophy in a New Key*, 3rd ed. (Cambridge: Harvard University Press, 1973); T. L. Short, *Peirce's Theory of Signs* (Cambridge: Cambridge University Press, 2007); and Margolis, “Rethinking Peirce's Fallibilism.”

³³ See, for instance, Alva Noë, *Action in Perception* (Cambridge: MIT Press, 2006).

Cassirer obviously goes beyond the most canonical of Kant's constraints: the inclusive and changeless categories said to govern all sensuous intuition. In doing that, Cassirer signals the cognate effect of additional creative possibilities bearing on categorial "thought" among inquiries involving other of Cassirer's "symbolic forms." But none of this would have been conceivable if we lacked a sufficiently enriched account of what the "I" signifies: the evolving, self-inventive self or agent of any human inquiry whatsoever – a figure closer therefore to Hegel than to Kant but distinct from both. Cassirer's innovation ranges without let or reference to Kantian systematicity over any sustainable inquiry. Cassirer obviously breaks with Kant's *transcendentalism* but is loath to admit the fact explicitly. There's the breakthrough that topples Kant's apriorism and signals the insuperably historicized cast of the *a posteriori* second-order conjectures of what a science or objective inquiry is or must be.

VI

I must make my amends now. I've been as careless as Hume and Kant about the nature of the human self. Well, perhaps not altogether, since we've gained important ground against their utterly disastrous and inexplicable retreat. Because it's obvious that neither Hume nor Kant – nor Sellars, to join the centuries in a pointed way – could possibly account for their respective analyses of mind and world without admitting *the full powers of a competent human agent*. (I have yet to explain what such an "agent" must be like. I shall leave that question largely unanswered here. The answer requires a fresh beginning.)

Indeed, very nearly the entire question has been reopened in recent years in a distinctly courageous (entirely serendipitous) way by a young Australian philosopher, David Chalmers. Chalmers confronts in great detail the entire reductionist and eliminativist brotherhood (to which he would otherwise happily belong but for the unresolved puzzle of "consciousness"). He intends much the same puzzle that was posed more than thirty years ago by Thomas Nagel with regard to the "subjective aspect" of phenomenological "experience," which has never been satisfactorily explained.²⁸ Nagel was right of course; but Nagel never cast his own challenge in a way that might have precluded dualism. That, ultimately, is what I think we must recover. Dualism is itself a relatively innocent (but blind) strategy for impoverishing our conception of the self along reductionistic lines.

Chalmers's question is not quite the deeper question Hume, Kant, and even Sellars signal (without answering). Nevertheless, what he offers mobilizes the entire literature of the philosophy of mind in a more perspicuous way than the 18th-century or the usual 20th-century theories of mind (which Chalmers effectively dismantles) had ever managed to afford. I mean: the question of what a person or self (a subject or agent) must be, in any sense that concedes a "first-person" or phenomenological dimension to human life.

Chalmers broaches no more than the question of consciousness biologically confined, and he does so only against the backdrop of an otherwise apparently adequate reductionism. But in pursuing the failure of reductionism *there*, he inadvertently opens the entire

³⁴ Herbert Feigl seems to have introduced the notion of "nomological danglers" into the literature of the philosophy of mind; Sellars's danglers are more radical. See Herbert Feigl, *The "Mental" and the "Physical": The Essay and a Postscript* (Minneapolis: University of Minnesota Press, 1958, 1967).

³⁵ Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1953), §621.

reductive undertaking to a measure of scrutiny that can no longer preclude the culturally “penetrated” (culturally transformed) powers of the artifactual selves that we are. The Kantian origins of positivism made it possible for positivism’s late Anglo-American progeny to ignore the post-Kantian and Hegelian complication for a very long time. But now that that company has itself begun to consider incorporating Hegel’s distinctive resources, Chalmers’s appealing challenge cannot fail to resonate with the latent lessons recovered for example by reviewing Cassirer’s deeper challenge. A similar confrontation addressed to the standard account of the methodology of the sciences in general featuring reductionism, the analysis of causality, and a commitment to exceptionless covering laws, is now in the process of being revived and deepened (in tandem with the first) by what amounts to a second reading of Kuhn’s *Structure of Scientific Revolutions*. Thomas Nagel’s well-known anticipation of Chalmers’s complaint required Chalmers’s added voice in order to be effective both in the setting of the philosophy of mind and in the philosophy of science – and in our reflections about culture and history.

There’s the matter Hume and Kant and Sellars, in addition to the champions of scientism and the laxer champions of “hermeneutics” in Charles Taylor’s sense, have largely ignored. The fact is, a “reformed” reductionist like Chalmers (reformed in rather the same sense in which an alcoholic is “reformed”) holds precisely the same minimal view about the *functional* powers of human agents (for instance, “the reportability of mental states,” “the ability of a system to access its own internal states,” “the ability to discriminate, categorize, and react to environmental stimuli,” “the integration of information by a cognitive system,” “the focus of attention,” “the deliberate control of behavior”) that Charles Taylor, as the devoted opponent of reductionism and scientism that he is, would surely be prepared to avow.²⁹ But they would intend completely different readings of the questions broached: both would restrict themselves to “biology” of course, but Taylor would feature hermeneutic concerns and Chalmers would regard his own examples as yielding to some form of reductionism or computationalism.

It’s here that Chalmers falters – just where Taylor misreads the persistent eligibility of the reductive option: for Taylor believes the experiential or phenomenological complication of the “functional” powers of persons instantly entails the defeat of reductionism, whereas Chalmers believes the “functional” aspect of the powers (just enumerated) *can* be accounted for reductively. It’s only the “additional” presence of consciousness or experience (in the biological sense) that threatens to defeat reductionism! Chalmers fails – utterly and tellingly – because he *separates* what he calls “the hard problem” of consciousness from “the easy problem” of all the functional achievements (according to his lights) of the “I’s” mode of functioning. But of course he means by this that the easy problems do not in principle require consciousness in his sense though they can tolerate its presence. Whereas the essential point remains that the artifactual powers of the “I” cannot be captured at all by any purely biological model of consciousness but requires an account of the hybrid form of reflexive intelligence (“self-consciousness,” if you like) that dawns only with the mastery of language and its accommodating culture. As I say, the “I” is thoroughly artifactual, the continually ramified achievement of the power of enlanguaged self-reference; whereas “mind” is initially emergent in the biological sense: in the human case, of course, the “I” is inseparable from the culturally “penetrated” (transformed, second-natured, hybrid) powers of what is originally given (biologically) as mind.

Both Chalmers and Taylor go terribly wrong. Both miss the deeper issue: the discovery that what belongs to the subjective experience and agency of persons is already inextricably implicated in *their functional ability to hear and comprehend speech, to see a Vermeer interior, to produce Michelangelo's David, to test a mathematized hypothesis of physical theory with regard to what may be sensorily perceived* (in the way Cassirer stipulates). There's the heart of the matter, the still-elusive link that ensures the common ground that the arts and sciences share. The functional treatment of computers need not be the same as the functional treatment of selves. Chalmers puts the point memorably – and innocently – thus: the easy problems [of consciousness] are easy precisely because they concern the explanation of cognitive *abilities and functions*. To explain a cognitive function, we need only specify a mechanism that can perform the function. The methods of cognitive science are well-suited for this sort of explanation, and so are well-suited to the easy problems of consciousness. By contrast, the hard problem is hard precisely because it is not a problem about the performance [or “causal role”] of functions [“in the production of behavior that a system might perform”]. The problem persists even when the performance of all the relevant functions is explained.³⁰

But that is also what is being contested! For if Chalmers's problem of consciousness remains unresolved – that is, if the consciousness *implicated in* the human version of the “functions” he concedes is not demonstrably reducible, if function and consciousness cannot be disjoined in the central human cases – then selves and persons (and *their functions*) cannot be reduced either. Until we find a satisfactory resolution, the theory of the human and natural sciences and of the sciences and arts will never escape the constructivist alternative I've been hinting at. Chalmers embraces as a problem what Sellars appears to dismiss in the way of a solution! But he fails to see that in advancing his worry about “consciousness” he himself has, by conceding Nagel's original challenge about reflexive experience, inadvertently opened the inquiry to include the enlanguaged and encultured complexities of “self-consciousness.” Seen this way, it's an open question whether, in the human paradigm, Chalmers's reductive solution of the “easy problems of consciousness” is not generally dependent on the solution to the “hard problem”; hence, that the extension of the reductive solution to so-called easy problems without invoking the hard problem of consciousness at all remains a benign concession condoned by the intransigence of the other.

There are, of course, different models of phenomenological perception to choose from – and, indeed, others that will surely appear in due time. But we can already see that it would be advantageous if our phenomenology avoided every form of cognitive privilege and foundational presumption (of the sort one finds in Descartes and Locke, in the original *Protokolsätze* of the Positivists, in Husserl's account of transcendental phenomenology,³¹ and of course in Kant's apriorism) without abandoning the artifactual nature of phenomenology itself. For phenomenology essentially features the conditions of reportage vis-à-vis what is “given” in any mode of first-person experience; and *that* cannot fail to be culturally penetrated.

There's the fatal limitation of any Kantian approach to phenomenology: Kant insists on the passivity of sensuous intuition. Hence, we must make provision, phenomenologically,

for the transformed (or *gebildet*) or culturally “penetrated” nature of the “given” (as sound is “perfused” by meaning when transformed into speech). This would account for a pertinent form of conceptual “adequation” between world and mind – within the working space of our artifactual modes of cognition (or interpretation or appreciation or the like). It also confirms a rough similarity between the point of Wittgenstein’s original question about action and bodily movement and Husserl’s question about the “natural” and what is putatively “pure” in the phenomenological sense: except that Wittgenstein eschews all privilege and Husserl’s solution relies on privilege.

You may then also anticipate that to admit the *gebildet*, hybrid, “second-natured” nature of our particular mode of life suggests a generic formula for specifying any and all versions of encultured mind (within the sciences or other possible “symbolic forms” of cultural engagement), according to which what counts as information, meaning, semiotic import of any kind can be effectively accessed by apt agents and selves whose brains and bodies and enviroing world are already coded for such feats. Notably, both Peirce’s theory of signs and Cassirer’s semiotics³² need not disallow the accessing of information outside the brain – as among computers, biochemical processes, causally affective actions, suitably constructed artifacts such as musical performances and paintings.³³

Considerations of these sorts begin to impose conceptual liens on what to suggest as a reasonable theory of the human self without actually privileging any particular “metaphysics.” Embracing the minimal lesson of Kant’s Copernican turn (corrected along Hegelian lines), as that has been construed from Kant’s time to ours, we may indeed acknowledge that there are always a number of promising – potentially incompatible – ways of modeling the human “subject” or “agent” with an eye to providing an adequate theory of culture or of the unity of the natural and human sciences or the sciences and the arts. Whatever has been adduced thus far is meant to be as neutral as possible relative to arguments already mounted. Seen this way, we are, I would say, plainly pressed in a direction that has not yet been seriously considered among the most influential movements of Western metaphysics, nor satisfactorily defined by other movements known to be sympathetic to conceptions like the one I have in mind.

Here an obvious (entirely provisional) economy suggests itself. For one thing, eliminativism has never come to terms with the “dangling” status of persons and their intentions,³⁴ which (as reported) Sellars is content to add (rhetorically) to what he calls his “scientific image”: I say, dismiss the option as completely inoperative – for instance, for failing to accommodate Cassirer’s insight about the import of the shift in physics from Galileo and Newton to Helmholtz and Hertz. For another, reductionism may be put on indefinite hold, inasmuch as no satisfactory exemplar has ever been formulated, though reductionism itself need never disallow the central mass of anthropocentric distinctions (including consciousness) normally invoked in speaking of selves and persons: I say, put reductionism on hold until it can bring a promising materialist (or functional) paraphrase to the table. For a third, dualism is a philosophical scandal or a seriously inchoate intuition; in fact, failed reductionisms often appear as disappointing dualisms. I say, dismiss dualism in

the same spirit in which it would be right to dismiss eliminativism. If, with me, you go this far, then I suggest there is no promising strategy left except a constructivism that accords with what I've been calling internal and external *Bildung* – particularly an account in accord with the latter, since the latter is meant to explain the sense in which the former is operative at all. If you allow this much, then it should be clear as well that our theories are bound to be guided by the Darwinian discovery. I see no better advice than a word in favor of combining these last reflections with the positive gains due to admitting the primacy of phenomenology and the penetration of the cultural.

I have not yet said what we should mean by the terms “person” and “culture” – or by “a person” and “a culture”; actually I've kept myself from answering. Because, for one thing, the argument I've just laid out is something of a prolegomenon to any fruitful answer of the definitional sort; and for another, because what we learn informally from what we take to be salient in that prolegomenon is likely to be more important in our grasp of a larger world of relations than whatever precision we may bring to theories and definitions grounded in a contingent and shallow intuition. I shouldn't want any quarrel about the limited merit of any specific posit of mine (regarding the definition of “person”) to detract from the force of an ampler line of reasoning (the one just sketched) that seems (to me at least) to lead irresistibly in the direction of:

(a) construing *persons* or selves as artifactual transforms of the members of *Homo sapiens*, constructed or constituted by acquiring certain fluent *sui generis* powers unique to themselves: speech for instance; self-reference; enlanguaged thought, perception, feeling, intention, judgment, interpretation, choice, commitment, creativity, and learning;

(b) construing *culture* as a historically specific ensemble of processes of internal and external *Bildung*: artifactually formed, collectively shared and transmitted among cohorts of aggregated persons, through the effective force of which new generations of persons are first formed and continually transformed, and through which, also, societies are kept viable, under the condition and effect of their self-transforming histories.

These are the touchstones of my theory, valid (if they are valid at all) in virtue of one very large conviction: namely, that the whole of the specifically human world is the upshot of the transformation of our biological talents – the incarnation of meaning in sound for instance, as in transforming sound into speech (internal *Bildung*) – a process first made possible, evolvingly, by the *sui generis* emergence from animal sources of true language and its attendant cultural powers (external *Bildung*). Whatever of the human world may then rightly claim independent standing of any kind – speech, history, behavior, institutions, art and technology, archives of any sort – will be seen to be hybrid artifacts “uttered” by individual and aggregated persons, intrinsically open to interpretation and altered meaning as a result of having been produced or interpreted or submitted to interpretation under the evolving powers of human inquiry.

But what does that entail, you ask, when contrasted with dualism, reductionism, and eliminativism? Minimally, the only possible answer that I can think of is some *sui generis* form of emergentism: *artifactual*, for one thing, because it depends on the artifactual incipience and evolution of true language, *not* Darwinian though it requires the overlapping contributions of biological and cultural evolution, where the second cannot be reduced to the first and is essentially different from it; and, for another, *indissolubly incarnate* in biological things (preeminently, our self-constituting selves and their second-natured aptitudes) or incarnate in mere physical things, as what selves can utter, do, create, or produce (as, say, speech, actions, paintings, machines, technologies, histories, traditions

and the like), hybrid transforms of natural things so constituted by virtue of the enlanguaged and encultured activities of apt selves.

If, now, we turn back to Wittgenstein's prescient question: "what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm"³⁵ – a question Wittgenstein does not explicitly answer – we ourselves may now venture a distinctly worthwhile answer by way of construing "my raising my arm" as a metonym for encultured "utterances" of any and every kind. For one thing, there *can* be nothing "left over" since subtraction here, read literally, would entail an unacceptable dualism. *Per contra*, if the question is meant to query what is it that transforms a physical movement into an action (or a sound into a word, or a painted canvas into a painting), the answer is that it is rightly construed as what is duly "uttered" by the "utterance" of some person or persons – what, minimally, remains as culturally significant when the subtraction is completed.

Rearing an infant in a society of apt persons is, effectively, uttering a new person – transforming the infant into an apt agent within its home society. I utter a formal garden and a curse and a murder and a poem in different ways, but they all implicate the same "internalist" form of causality (if you agree to call utterance or agency a kind of causality). Our world is becoming increasingly artifactual, so that physical nature itself is abstractly constituted by a sort of subtraction akin to what Wittgenstein has in mind; that is, nature is so designated by the subtractive efforts of an artifactual science, even if it is not itself ught to have been culturally uttered or constituted as such.

That explains at a stroke the difference between the kinds of "objectivity" accorded the physical and the human sciences and arts and the sense in which inquiries and activities of every human kind implicate the role of persons in essentially the same way. We may then answer Wittgenstein thus: whatever might otherwise be deemed to be a "natural thing" (a thing that may be rightly described and explained in terms of natural causes) is judged, when deemed to belong intrinsically to the world of human culture, to have been transformed or penetrated so that it is rightly said to possess in a public way – *its* "nature" being adequated to its possessing in some objectively determinable way – indissolubly incarnate attributes of cultural significance and signification (expressive, representational, signific, symbolic, semiotic, linguistic, intentional, institutional, normative, historical). Which, therefore, profoundly alters our conception of objectivity, a valid science, causality, the causal closure of the physical, the laws of nature, scientific inquiries beyond the scope of causal explanation among the physical sciences, truth, universality, material necessity, the rigor and nature of a human science and cognate disciplines, and the very idea of a science of interpretation.

This much, I think, is relatively uncontroversial to all philosophies that eschew eliminativism and dualism or that oblige reductionism to provide a reasonably promising exemplar of a science or discipline of its own choosing before reductionism itself may claim more than second-class citizenship. These are philosophies prepared to enrich our theory of selves or persons in ways that compare favorably with what is usually taken to be the normal run of competences of the apt members of a society of persons. Beyond all that, questions turn quarrelsome as they become more detailed – often therefore more interesting. In the meanwhile, we begin to see the general lineaments of a conception of the unity of the natural and the human sciences, and of the arts and sciences, that goes entirely contrary to the reductive motivation of the canonical accounts. It's the ubiquity of the self

that entails the unity of all the forms of human inquiry and agency. And the self is itself an artifact of the human form of life produced at long last by the processes of external *Bildung* now ubiquitously regularized in the historied forms of internal *Bildung*.