Acta Anal (2010) 25:413–434
DOI 10.1007/s12136-010-0095-8

# Mind Stuffed with Red Herrings: Why William James' Critique of the Mind-Stuff Theory Does not Substantiate a Combination Problem for Panpsychism

**Itay Shani**

Received: 16 November 2009 / Accepted: 4 March 2010 / Published online: 15 April 2010


**Abstract** There is a famous passage in chapter six of James' *Principles of Psychology* whose import, many believe, deals a devastating blow to the explanatory aspirations of panpsychism. In the present paper I take a close look at James' argument, as well as at the claim that it underlies a powerful critique of panpsychism. Apart from the fact that the argument was never aimed at panpsychism as such, I show that it rests on highly problematic assumptions which, if followed to their logical consequences, are just as inedible to contemporary critics of panpsychism as they are to its present-day supporters. Hence, a naïve employment of the argument, as a critique leveled by physicalists against panpsychism, is counterproductive and even self-defeating. After examining the metaphysical shortcomings undermining James' position (as well as the hasty "refutations" of panpsychism based on it), I conclude with some reflections on what needs to be done in order to obtain a better perspective regarding the explanatory prospects of panpsychism as an alternative approach to mainstream physicalism in the study of conscious phenomena.

**Keywords** Aggregates · Cohesion · The *combination* problem · Mild emergence · Mind-stuff theory · Radical emergence · Unities

## 1 The Tu Quoque Argument Against New-Wave Panexperientialism

One of the consequences of the surge of interest in consciousness over the past twenty years or so is the revival of the idea that experience may well be a fundamental feature of reality. While committed defenders of panpsychism,

"If evolution is to work smoothly, consciousness in some shape must have been present at the very origin of things". William James, *Principles of Psychology* Vol. I, 149.

I. Shani (✉)
Department of Philosophy, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa
e-mail: ishani479@hotmail.com

panexperientialism (Griffin 1998), panprotopsychism (Chalmers 2002), 'real physicalism' (Strawson 2006), 'o-physicalism' (Stoljar 2001) or any other name by which the rose is being designated are still a minority in professional circles, and while the view they defend still counts as heterodoxy, the "idea" is on the ascent, steadfastly gaining credibility as a serious challenger of contemporary mainstream physicalism.[1]

In challenging mainstream physicalism, one of the polemical fortes of this "new-wave panexperientialism" consists in pointing out the difficulties inherent in the doctrine of *radical emergence* (see, e.g., de Quincy 2002, 221; Strawson 2006), namely, in the idea that phenomenal consciousness emerges from an utterly insentient physical background. Neo-panexperientialists maintain that careful scrutiny reveals that such radical emergence—a complete breach of ontological continuity, the sudden appearance on the scene of something wholly unprecedented, devoid of any primitive forerunners, and of which no traces can be detected prior to its ultimate surge—is hopelessly unintelligible.

A major credential of the panexperientialist alternative lies in its promise to alleviate this burden of unintelligibility. Sentience does not emerge out of a wholly insentient physical background; rather, it has been there from the beginning, albeit in a spectacularly scaled-down form, and its evolution towards the powerful forms of phenomenal consciousness known to us from our own experience is but an expression of a general cosmological nisus towards increased organizational complexity. Complex phenomena emerge out of interactions between phenomena of lesser complexity but this type of process is one of *mild emergence*, where the latter emergents can be intelligibly traced to their humble origins (even if they cannot be reliably predicted). Ditto, the idea goes, with experience: powerful forms of sentience arise out of the integration of, and interactions between, systems endowed with lesser, yet non-nugatory degrees of sentience.

Yet this argument is being subjected to a counterargument, a *tu quoque* argument. Critics of panpsychism argue that the emergence of sophisticated forms of experience out of primitive forms of experience is just as mysterious as the emergence of experience out of wholly insentient physical antecedents (see, e.g., Carruthers and Schechter 2006; Goff 2006; Van Cleve 1990). If so, if the mild emergence favored by panpsychists is as arbitrary as the radical emergence of orthodox emergentist physicalism, then new-wave panexperientialism looses its epistemic edge over the physicalist orthodoxy, with little remaining to motivate the radical ontological revision it espouses.

The apparent unintelligibility of the emergence of complex experience out of the integration of, and interaction between, elements endowed with simpler forms of experience is known as the *combination* problem. William Seager, who coined the term (Seager 1995) made it clear that the locus classicus of this problem is William James' critique of the *mind-stuff* theory (James 1890/1950, vol. I, chap. 6; but see Skrbina 2006 for some comments on the more remote history of this problem), and it

[1] In what follows I use 'panpsychism' and 'panexperientialism' freely and interchangeably. As far as I can see, none of what I argue for depends crucially on the precise labeling of this metaphysical alternative to standard physicalism. If the reader feels more comfortable with a particular labeling (e.g., 'panprotopsychism') he (or she) is welcome to do so to his (or hers) heart's content.

is customary among critics of panpsychism to use James' pointed critique of the mind-stuff theory as effective weaponry in the debate (two recent examples are the aforementioned papers by Carruthers and Schechter, and by Goff).

However, a careful look at the conceptual terrain pertaining to the combination problem, and in particular at James' actual argument, reveals that as a general *tu quoque* argument against panpsychism the effectiveness of James' argument is seriously limited. No doubt, the panpschist's claim to have the edge over the radical emergentist in explaining the emergence of evolutionarily advanced forms of consciousness needs to be justified. Hence, one could argue that as long as it remains unclear precisely how a panpsychist ontology dispels the dark clouds surrounding the emergence of rich and complex consciousness out of interactions between elements whose own phenomenal attributes are exponentially simpler there lurks the concern that its promise of transcending the blind-alley of epistemic arbitrariness may not be realized.

Arguably, one could even call this problem 'the combination problem' provided that one remains vague enough, for all we are actually saying is that we need an account of how the hypothesis that even the tiniest material systems are imbued with sentience plays a vital role in the explanation of evolutionarily advanced forms of consciousness. This, of course, is a very general requirement: it is perfectly possible that such an account will *not* be predicated on the assumption that a sophisticated consciousness is literally composed of a multitude of simple consciousnesses. But, that is not how the combination problem is presented in the literature, for the standard meaning given to this problem is considerably narrower, implying that it consists of the puzzle of how elementary units of experience can *merge* into higher forms of experience (Seager 1995). Worse still, as I show below, this merger, or combination, is often believed to be strictly aggregative, with the emerging complex experience being nothing but the *sum* of constitutive rudimentary experiences. This is precisely the idea behind the mind-stuff theory attacked by James. Hence, a major weakness of the tu quoque argument against panpsychism is that the combination problem applies only to a certain *variant* of panpsychism, one among many generic forms which the idea can assume (cf. Skrbina 2006).

Moreover, the point is not merely logical. Rather, it is that nowadays we will be hard pressed to find anyone who defends a version of panpsychism as atomistic and aggregative as the mind-stuff theory. Indeed, what is so striking about the mind-stuff theory is that its atomism and compositional aggregativity excludes the possibility of emergence of *any* sort, let alone the emergence of powerful forms of consciousness. It is therefore not surprising that few, if any, of the contemporary sympathizers of panpsychism find such a framework hospitable. What is surprising, or better odd, is the extent to which this red herring is used so uncritically by able defenders of standard emergent physicalism.

Finally, the weight assigned to this argument is even more puzzling given that all these problems are easily detected in James' own writing. Reading James' oft-cited chapter it is clear that his critique of the mind-stuff theory was *not* an all-out attack on panpsychism but was aimed at a theory that, at best, is a highly specific version of panpsychism and which, in some of its manifestations, may not even qualify as such. Moreover, it is also clear that James' discontent with the mind-stuff theory was primarily on account of its atomism and compositional aggregativity, even if James

himself was unable to fully go beyond these metaphysical commitments (cf. Henle 1990).

In what follows I take a closer look at the scope and force of James' argument with the aim in mind to unravel what it does, and what it does not tell us about the feasibility of panpsychism as a general framework within which to account for the evolution of consciousness. It will then be argued that while James' argument against the compositional additivity of experience is cogent the fact that experiences do not sum need not deter sympathizers of panpsychism and of the doctrine of mild emergence since direct summands are the *antithesis* of genuine emergence, of any sort. I conclude with some reflections on the current state of the debate between mainstream physicalism and new-wave panexeprientialism in the aftermath of the realization that James' argument cannot be put to use in the sweeping manner which many have intended it to be.

## 2 A brief Historical Aside on James' Critique of the Mind-Stuff Theory

Although my motivation for exploring James' critique of the mind-stuff theory (MST) is conceptual rather than historical, and although I am concerned with only one specific argument hailed by James against this theory (namely, with his attack on "the self-compounding of mental facts"; see section three below), it is advisable to begin our discussion with a brief historical supplement else we fall prey to oversimplifications, misrepresentations and even downright mistakes.

The first thing which must be born in mind is that James himself did *not* consider his critique of MST, and in particular of the doctrine of the self-compounding of mental facts, as constituting a general argument against panpsychism. That this is so is clear from his stance towards the evolution of consciousness. Comparing the idea of radical evolutionary emergence with the view that consciousness, however rudimentary, is ontologically fundamental, James' sympathy lies clearly with the latter position. Thus, he proclaims rather unequivocally that:

> [C]onsciousness, however small, is an illegitimate birth in any philosophy that starts without it, and yet professes to explain all facts by continuous evolution. *If evolution is to work smoothly, consciousness in some shape must have been present at the very origins of things*" (James 1890/1950, 149).

Doubtless, this assertion could not have been made by an avid crusader against panpsychism.[2] Moreover, latter on in this chapter (chapter VI of the *Principles of Psychology*) James openly acknowledges the fact that some varieties of panpsychism, in particular Leibnizian monadism, escape the critique he levels against MST. That James was critical of the monadic theory of the mind for independent reasons is besides the issue, since the point I wish to make is that James never intended his critique of MST as an all-out attack on panpsychism.[3]

[2] James' reluctant sympathy to panpsychism turned into full endorsement later on in his intellectual career (see Skrbina 2006).

[3] Although James disapproves of the monadic theory of the mind as being too abstract and detached from experience, he concludes his discussion of it on a more positive note, asserting that "a theory which Leibnitz, Herbart, and Lotze have taken under their protection must have some sort of a destiny" (queryJames 1890/1950, 180).

Finally, and again in support of the contention that the demise of panpsychism was never the goal of James' critique of MST, it is worth noticing that the theory criticized by James was a concrete theory popular at the time, of which panpsychism was only one constitutive element.[4] Interestingly it is the other elements of the theory, rather than the commitment to panpsychism per se, which James targets for criticism. Thus, James criticizes at length the *elementarism* of MST (namely, its commitment to a compositional atomism wherein complex mental processes are composed of simple mental elements all they way down to the ultimate mental "atoms") as well as the strict *parallelism* it postulates between mental structures and physical structures (viz. the contention that each level of physical complexity is accompanied by a corresponding level of mental complexity: physical atoms are accompanied by mental "atoms", physical molecules by mental "molecules", and so on).[5] But while these principles are subjected to harsh criticism there is never, in the whole chapter, a direct attack on the panpsychist core of MST, that is, on the "hylozoist" assumption that consciousness is aboriginal.[6,7]

Clearly, then, a critique of MST, however devastating, is not necessarily a devastating critique of panpsychism. Of course, the situation would become dire for panpsychists in the eventuality that the fire and brimstone James throws at the mind-stuff theory were, for some reason, applicable across a wide spectrum of panpsychist variants, including the most promising variants available at the present. To see whether this is the case, we need to take a closer look at James' most celebrated attack on MST.

## 3 James' Master Argument Against the "Self-compounding of Mental Facts"

James' central objection to MST consists in his celebrated argument against the *self-compounding of mental facts* (James 1890/1950, 158–162), which argument forms the backbone of what many identify today as the *combination* problem for panpsychism. Simply put, James argues that the idea of mental units 'compounding with themselves', or 'integrating' is unintelligible. In a famous passage, which proponents of the tu quoque argument like to cite, he proclaims that a hundred separate elementary feelings would never coalesce into a unified complex feeling and ditto, that if we take a sentence of a dozen words, and give 12 persons a word each to think about, "nowhere will there be a consciousness of the whole sentence"

[4] The term 'mind-stuff' is due to W.K. Clifford (1878); James also associates the doctrine with the names of Fechner and Haeckel, while Clifford himself ascribes "the first statement of the doctrine" to Wundt (Clifford 1878, 67). As will be seen in the next section, James' critique of MST was influenced by an earlier critique due to Royce (1881).

[5] For more on James' critique of these conceptual elements of MST see Henle (1990).

[6] *Hylozoism*, a term first used by the 17th-century Cambridge Platonist Ralph Cudworth, is the view that life and mind are inherent in all matter. James uses this term here as a synonymous with 'panpsychism'.

[7] Noticeably, James' arch-rival throughout much of the chapter (in particular pp. 151–158) is Herbert Spencer, even though Spencer himself was not even a true panpsychist! (He believed the simplest forms of conscious experience emerge with certain primitive neural ganglia). What makes Spencer an apposite target for James' attack (apart from what seems like a personal disliking…) is the fact that his writing nicely illustrates the elementarism and parallelism which James so vividly resists (needless to say, Spencer's parallelism did not stretch beyond the mental equivalents of the lowest level of the aforementioned primitive neural ganglia).

(James 1890/1950, 160) to be found. In short, "the private minds [viz. micro-experiences] do not agglomerate into a higher compound mind [a macro experience]" (Ibid).

Advocates of the tu quoque argument take this oft cited passage to demonstrate that there is a *combination* problem for panpsychism whose gravity is on par with the *generation* problem plaguing mainstream physicalist theories of consciousness (i.e., those based on the idea of radical emergence).[8] In the words of Philip Goff:

> [t]he emergence of novel macroexperiential properties from the coming together of microexperiential properties is as brute and miraculous as the emergence of macroexperiential properties from non-experiential properties. (Goff 2006, 54).

As was mentioned in the opening section, if this tu quoque charge is well-founded than it counts heavily against the explanatory aspirations of panpsychism. But how serious the charge really is? Since the apparent unintelligibility of the "self-compounding of mental facts" (SCMF, for short) forms the core of the tu quoque assault on panpsychism and is the reason why James is regularly cited in present-day debates on the subject, it is worth our while to examine precisely why James concluded that such integration is unintelligible. Only by tracing the internal logic of the argument will we be able to assess its strength and to adjudicate on whether or not it ought to weigh heavily in the present context.

In attempting to unearth the underlying logic of James' argument, two related points stand out, calling for immediate attention:

(I) James' argument against SCMF is an argument against the existence of *intrinsically* integrated mental wholes.
(II) The argument is but a special application of a *general* argument against the very possibility of intrinsically organized wholes of *any* sort.

The first point must be borne in mind if we are to understand how James concludes that the idea of mental units "'compounding with themselves' or 'integrating'" is "logically unintelligible" (James 1890/1950, 158). The second point is even more significant in that it reveals the core weakness of the argument, namely, the fact that it comes at the price of a categorical denial of emergent organized complexity. Following James' reasoning, it will become clear that his argument is framed within a framework which makes the whole gamut of emergent causal unities illicit. Ironically, the absurdity of this eliminativist result seem to have been lost on those eager to employ James' argument at the service of the present debate between physicalists and panpsychists.

To begin to appreciate the general outlines of James' stance and his reasons for concluding that SCMF is unintelligible there is, I believe, no better way than to quote a whole paragraph from Josiah Royce whose '*"Mind-Stuff" and Reality*' (Royce 1881) was an earlier attack on MST which clearly influenced James. Apart

[8] Like the 'combination problem', the 'generation problem' is a term coined by Seager (1995). Essentially, the problem is none other than Chalmers' celebrated 'hard problem of consciousness' (Chalmers 1995), namely, the problem of explaining how process-configurations of insentient material elements may culminate in phenomenal experience.

from the fact that James himself approvingly quotes the last four sentences of this paragraph, it will be seen that the paragraph contains three theses which are key in James refutation of SCMF.

> In nature it often happens that a manifold of distinct parts results in a unity that is not a mere sum. So every organism, so every chemical compound, exhibits properties qualitatively different from the properties of the constituent parts. But how are these properties of the compound manifested? Only in the behavior of the manifold towards the world external to itself. In itself a mass of parts, the whole behaves as one when it comes into relations with other things. The compound molecule is a sum of atoms. But in its behavior as a molecule towards other molecules it shows new qualities, and so is more than a mere aggregate. The organism is an aggregate of tissues. But in its behavior in the presence of the outer world it shows adaptation and an integration of parts, so that we call it one, not a mere aggregate. A mere combination is, when regarded in and for itself, never an organized whole. Aggregations are organized wholes only when they behave as such in the presence of other things. A statue is an aggregation of particles of marble; but as such it has no unity. For the spectator it is one; in and for itself it is an aggregate: just so, for the consciousness of an ant crawling over it, it may again appear a mere aggregate. No summing of parts can make a unity of a mass of discrete constituents, unless this unity exists for some other subject, not for the mass itself." (Royce 1881, 375–376; partially quoted on page 159 of the *Principles of Psychology*).

As mentioned above, this paragraph contains three theses which play a key role in James' critique of SCMF. These theses are:

(1) Aggregativity (AG): All complex natural systems are aggregates, the *agglomerative sums* of their proper parts.
(2) The non-unity of aggregates (NUA): Mere aggregates do not constitute intrinsically organized wholes (viz. aggregates aren't unities).
(3) The tertium quid principle (TQP): The only sense in which the elements of an aggregate may, and sometimes do, act as a unity is in exerting *combined influence on an external entity*, a third thing ("tertium quid", see James 1890/1950, 157).[9]

The endorsement of AG is clear enough from the assertion that molecules and organisms (which are mentioned here as exemplars of complex wholes) are "sums" and "aggregates", as well as from the more general contention that "in itself" a whole is nothing but a "mass of parts". Ditto, NUA is affirmed in such phrases as "no summing of parts can make a unity of a mass of discrete constituents" and "[a] mere combination [understood here as mere aggregation] is, when regarded in and for itself, never an organized whole". Finally, TQP is repeatedly announced throughout the paragraph and is stated most clearly when it is said that "aggregations are organized wholes only when they behave as such in the presence of other things".

Noticeably, there is a certain complementarity between NUA and TQP. The first thesis denies that aggregates are organized wholes in and of themselves, while the

[9] It may be observed that James ascribes this last thesis not only to Royce but also to Hermann Lotze who, according to him, "has set forth the truth of this law more clearly and copiously than any other writer" (queryJames 1890/1950, 159).

second thesis maintains that they act as organized wholes "only when it comes into relations with other things". In short, aggregates are not intrinsically unified and whatever unity one is tempted to find in them is wholly extrinsic, consisting entirely in their effects on other entities. But according to AG *all* compounds are aggregates, from which, in conjunction with TQP, it follows that:

> *All the 'combinations' which we actually know are* EFFECTS, *wrought by the units said to be 'combined'*, UPON SOME ENTITY OTHER THAN THEMSELVES" (James 1890/1950, 158; italics and bold letters in the original).

This appears to be James' most lethal weapon in his assault on SCMF. Let us label this thesis the 'generalized tertium quid principle', or GTQP. With GTQP in place James' refutation of SCMF is ready at hand. Simply put, if the unity of any given complex causal structure C is nothing but the combined effects of its proper parts on something external to C, then C lacks an *intrinsic* measure of coherence and unity. Consequently, to insist on the reality of intrinsically integrated unities would be to commit oneself to an idea which, in James' words, is "logically unintelligible" (Ibid). From this general argument against *self*-compounding (namely, against the reality of intrinsically organized wholes) the inadmissibility of self-compounding *mental* facts follows as a special case, or to quote James again "[w]here the elemental units are supposed to be feelings the case is no wise altered." (Ibid, 160).

This, then, is the abstract structure of the argument. Let us now see how the argument proceeds on a more concrete level. A careful selection of examples makes GTQP seem initially plausible. Thus, James maintains that each combination requires a *medium*, or *vehicle*, in order to materialize: a multitude of contractile muscles must all be attached to a single tendon in order to exert their joint dynamical forces; mechanical forces must operate on a body in order to combine into a diagonal vector sum; 'concord' and 'discord' are names for the combined effects of musical sounds on an external medium, the ear; and so on. "In other words", he concludes, "no possible number of entities…can sum *themselves* together" (James 1890/1950, 159); the sum itself exists only as an effect on an entity external to it.

As mentioned earlier, James concludes that the same moral applies to mental compositions: the 100 original feelings (viz. micro-experiences) may effect the creation of a 101st feeling (a macro-experience), but this new feeling, he argues, would be totally external to its predecessors—it would not contain them as intrinsic ingredients. The upshot, again, is that SCMF must be false.

No doubt one can appreciate the elegance of James' reasoning, but veracity is a different matter. In the next section I show that James' position runs into serious trouble; so serious that they ought to deter anyone from naively endorsing his conclusion.

## 4 Whither Emergent Complexity? Exploring Some Highly Unpalatable Consequences of James's Argument

The first noticeable problem with James' argument is the *empirical inadequacy* of GTQP. To put it bluntly, it simply isn't true that holistic integration is entirely, or even primarily, a matter of exerting influence on some tertium quid. Following Royce, James was led to conclude that no composite system is an organized whole

*in* itself, let alone *for* itself. In other words, no collection of micro-components $s_1 \ldots s_n$ could coalesce into an intrinsically unified, causally integrated whole. Consequently, on this view, the unity of a macro-system S, or better its appearance of unity, is restricted to the form of effects on, or representations in, some *other* systems S′, S″, etc. Clearly, then, this radically extrinsic account dismisses all allegiances to intrinsic measures of coherence and integrity. However, holistic unity, or *cohesion* as I shall refer to it here (following Collier, see, for example, Collier 1988; 2008; Collier and Muller 1998), is a characteristic fact of countless natural and artificial systems spanning the entire spectrum of physical complexity.

Roughly speaking, the cohesion of a system is a measure of its capacity to retain macro-level stability and coherence amid interferences caused either by interactions with things in its ambience, or by internal micro-level fluctuations (see Collier 1988; Collier and Muller 1998). A cohesive system is a system capable of maintaining macro-level invariance in the face of a significant range of destabilizing effects. The key to such macro-level stability lies in causal interactions—bonding relations—between the system's components, interactions which endow the system with an enduring organization, namely, which *make* it an organized whole. The point of mentioning cohesion is to stress the reality of an intrinsic dimension to holistic organization, irreducible to a purely extrinsic analysis.[10] To illustrate the point it is useful to touch briefly on a basic taxonomical distinction in complexity theory, namely, that between *energy-well stabilities* and *far-from-equilibrium* systems.[11]

Energy-well stabilities are systems with typically strong structural bonds whose stability is maintained near a state of equilibrium. The term 'energy-well' designates the fact that such systems lay low, as it were, at the ebb of a basin of attraction. Barring the application of a significant amount of energy to knock the system off its low-energy stable state its relatively strong local bonding relations ensure that sufficient number of the system's micro-components will remain arrested in their positions to maintain the integrity of the whole. As examples of energy-well configurations we may mention crystals, rocks, and planets, as well as artifacts such as marble statues and iron boxes.

Now, the point I wish to make is that the cohesion of such systems is an *intrinsic* feature of their (macro) organization. To be sure, the intrinsic organizational properties of a given system S affect the manner in which S interacts with its surroundings but the contention that S's unity is nothing more than its dispositions to affect *other* systems in certain orderly manners is a gross distortion. To take a very simple example, consider heavy water. There are several important respects, physical, chemical, and biological, in which heavy water differs from ordinary water in its capacity to affect a "tertium quid". For example, heavy water is a much more efficient moderator in nuclear thermal reactors, and a high concentration of the substance is toxic and disruptive of cell division, proving lethal to eukaryotic life forms (Routh 1965). However, such differences in the dispositional properties of

[10] Interestingly enough, James was rather skeptical with regard to the existence of any genuine bonds between material elements (see, for example, queryJames 1890/1950, 136).

[11] The distinction between energy-wells and far-from-equilibrium systems is sometimes captured under a different terminology. Thus, von Bertalanffy (1968) emphasized the distinction between 'closed' and 'open' systems, while in Prigogine's work the term 'dissipative structures' is often used as a synonym for 'far-from-equilibrium systems' while energy-wells are called 'equilibrium structures' (Prigogine 1978).

$H_2O$ and $D_2O$ are anchored in intrinsic structural differences. For example, the disruptive effects of high-concentrations of heavy water on cellular reactions are due to the fact that deuterium manifests stronger structural bonds than regular hydrogen. In short, what sets a $D_2O$ molecule apart from an $H_2O$ molecule as a source of potential effects on its environment is the fact that the two differ in their internal constitution and organization—simply put: they are different types of cohesive organized wholes.

The existence of an intrinsic dimension to holistic organization becomes even more conspicuous when we consider systems whose cohesiveness is maintained far from thermodynamic equilibrium. In contrast with energy-well stabilities, far from equilibrium systems are much more sensitive to change: their chemical bonds are much weaker, and they face a strong pull towards devolving into equilibrium and disintegration (Bickhard and Campbell 2003; Nicolis and Prigogine 1989). Under such conditions, prolonged systemic cohesiveness depends on a constant exchange of matter and energy with the environment and on the capacity to tunnel acquired energy at the service of a continuous maintenance of the system's stability.

This is true even with regard to inanimate systems—e.g., chemical clocks, vortices, and storm systems, to mention but a few (Capra 1996; Kauffman 1995; Prigogine and Stengers 1984). However, when it comes to the biological, psychological, and socio-cultural spheres the dependency of cohesiveness on active self-maintenance reaches spectacular new levels: For in living systems prolonged cohesiveness necessitates a coordinated and flexible (context-sensitive) control over the processes which underlie the system's viability. In other words, the very capacity of the system to maintain itself through interactions with its environment is being monitored, modulated, and maintained. In the language of complexity theory such recursive, reflexive, and proactive self-maintenance is often referred to as '*autonomy*'.[12] At its very basic level, for example in microscopic organisms, autonomy may consist of nothing more than some rather elementary sensing capacities, feedback and feedforward processes, metabolic processing, boundary maintenance, and motor capabilities. But, as we climb up the evolutionary ladder it gradually scales up to cover the whole gamut of motivation, emotion, imagery, and cogitation, of intrapersonal and interpersonal signaling and communication, of memory, anticipation and planning, of consciousness, language, symbolization, creativity, conventions, institutions, and much more.[13] In other words, when we reach the biosphere and the noosphere we reach a *heightened platform of intrinsic integration*, a realm of goal-directedness, self-regulation, self-preservation, self-construction, self-realization, and self-transcendence, a domain of *value* and *purpose* in which cohesive wholes are not only organized *in* themselves but also *by* themselves and *for* themselves.[14]

[12] More extended discussions of the concept of 'autonomy' can be found, for example, in Bickhard 2000b; Christensen and Hooker 2000; Kauffman 2000, ch. 3; queryRuiz Mirazo and Moreno 2000; Shani 2006; and Várela 1979.

[13] For discussions of the hierarchical structure of autonomous organization, from biological to cognitive to socio-cultural phenomena, see, for example, Bickhard (1998, 2004), Jantsch (1976), querySteiner and Stewart (2009).

[14] On self-realization and self-transcendence see Jantsch (1976). On the normative dimension of autonomy see Barham (2004), Bickhard (2006), Christensen and Hooker (2000), Laszlo (1996).

The discussion above illustrates not only that stable forms of holistic organization are commonplace but also that stability is an *intrinsic* macro-property of such organizations. What turns a collection of micro-elements into a unified whole is not merely, or even primarily, the fact that when conjoined they tend to affect some external entity in certain definite manners; rather, it is the overall coherence and stability resulting from the interactions between the parts which constitutes systemic unity. Such unity, it may be stressed again, is intrinsic to the emergent whole—it can be reduced neither to an impact on, nor to a representation in, some tertium quid, the latter being a different system altogether. In short, James' contention that no compounds are intrinsically organized has little to command in its favor. Whatever scientific plausibility it may have had in James' own day must surely be vanquished at the present.

But beyond the empirical inadequacy of GTQP, James' argument is also rife with philosophical problems. Perhaps the most daunting one is that by relinquishing the possibility of intrinsically organized entities it winds up embracing the conclusion that emergent causal unities are nothing but an illusion. Thus, James announces that "the sum itself exists only *for a bystander* who happens to overlook the units and apprehend the sum as such" (James 1890/1950, 158–159). And as if making the point in such palpable terms is not provocative enough, he goes on to assert that

> Atoms of feeling cannot compose higher feelings, any more than atoms of matter can compose physical things! The 'things,' for a clear-headed atomistic evolutionist, are not. Nothing is but the everlasting atoms. When grouped in a certain way, *we* name them this 'thing' or that; but the thing we name has no existence out of our mind. (Ibid, 161).

It is crystal clear then, that James' refutation of SCMF is bought at the price of a general eliminativism vis-à-vis emergent organized wholes, a heavy price indeed! In fact, it is not too steep to suggest that such radical eliminativism constitutes a reductio ad absurdum of any theory that entails it.[15] At any rate, it surely is bad news for those who are using James' argument in the context of the present debate between physicalists and new-wave panexperientialists. As was made clear in the opening section, both parties to the debate presuppose the reality of emergent causal unities and, within it, of emergent mental phenomena. What is being disputed pertains to questions such as: Is radical emergence ever takes place? Is the emergence of complex mental properties an instance of mild, or of radical emergence? And, are there any clear epistemic benefits to the assumption that the emergence of complex mental properties is mild rather than radical? Thus, within the context of this controversy an appeal to James' celebrated critique of MST is *self-defeating*, for the eliminativism with which James winds up negates the basic presupposition, common to both parties, that there *are* emergent organized wholes.

I shall return to this significant result as well as to the reasons behind James' denial of emergence later, but, for the time being, let us mention a few more philosophical difficulties with James' position.

[15] In addition, the fact that the "everlasting atoms" are now known to be neither structurless, nor changeless, combined with the holistic import of quantum mechanics and of quantum field theories, further undermines the plausibility of the contention that organization is illusionary and that only structureless atoms are real (I touch on this subject in Shani 2007; see also Bickhard 2000a).

Yet another, related problem with GTQP is that it yields *epiphenomenalism*. Recall, that the thesis asserts that an aggregate of separate parts "behaves as one" only "when it comes into relation with other things" (Royce 1881, 375), but the problem is that the unity thereby created at the terminal point is no more than an *appearance* of unity, created by correlated independent external causes. This point is illustrated nicely by Collier and Muller (1998). Consider phenomena such as motion pictures, or a salvo of bullets. The size, orientation, duration, or intensity of such groupings, not to mention their apparent causal interactions, are not of their own making. Rather, they are really epiphenomena of underlying external causes (the cat in the cartoon does not really chase the mouse; it is merely made to look *as if* it does). By contrast, a swarm of bees, or a flock of birds are held together by genuine causal interactions among the participants; the structure thereby maintained is of their own making. Clearly, the "unity" sanctioned by James, and by Royce, belongs to the former, epiphenomenal sort. Indeed, given the exclusion of intrinsic cohesion epiphenomenalism is all but inevitable: the order created in a tertium quid C, as a result of the combined effects of A and B, cannot be a genuine emergent organization for such emergence has been banned; therefore, it can only be an illusion of an organized pattern, an epiphenomenal shadow of cohesive unity.[16]

Finally, I would like to draw attention to yet another untenable ingredient in James' account, namely, to his contention that the only location wherein genuine integration takes place is within one's soul. James advances this idea (though with some reluctance) as the solution to the problems raised in his critical discussion of MST. The soul, he argues, is the medium "upon which (to use our earlier phraseology) the manifold brain-processes *combine their effects*" (James 1890/1950, 181). As such, it is not only a tertium quid, but, apparently, the right kind of such third entity. Indeed, James held that this hypothesis offers a *solution* to the problem of the integration of mental facts, arguing that since the soul's "pulses of consciousness are unitary and integral affairs from the outset, we escape the absurdity of supposing feelings which exist separately and then 'fuse together' by themselves" (Ibid).

It is not difficult to see, however, that this route of escape is a rather dubious one. First, as James was well aware, it opens up the Pandora box of dualism and the problem of interaction (although James himself was of the opinion that the intricacy of the problem of mind-body interaction is no more difficult than the problem of causation *within* each separate domain—the physical, or the mental).[17] Second, James' claim that "[t]he separateness is in the brain-world… and the unity in the soul-world" (ibid) is nowhere justified. Rather, it is simply taken for granted. But given the sweeping assumption that intrinsic unity is nowhere to be found outside the boundaries of the soul the question naturally arises in virtue of what does the

[16] Noticeably, such epiphenomenalism is against the grain of James' ingenious defense of the causal efficacy of consciousness as expressed in particular in his critique of the *automaton theory* (in chapter 5 of the *Principles of Psychology*). As such, this result constitutes a problem for the overall internal logic of his position (nor would it be the first time such a problem occurs, for James was a man of many views and many positions, which, though ingenuous, are often hardly reconcilable with each other).

[17] See James, ibid. I suppose that in making this controversial claim James must have had Hume on his mind. Earlier in the *Principles of Psychology* he says that "[a]s in the night all cats are grey, so in the darkness of metaphysics all causes are obscure" (James 1890/1950, 137).

soul alone able to muster such intrinsic integrative powers? It is easy to take it for granted that the soul somehow manages to do the trick given the phenomenological observation that, as James puts it, the "pulses of consciousness are unitary", but the fact that such unity is manifested in conscious experience in no way *explains* how it is possible in the first place, and James does little to alleviate this burden of possibility.[18] In the absence of such an explanation, however, invoking the soul as a solution has, to use an old saying of Russell, all "the advantages of theft over honest toil" (Russell 1919, 71).

In sum, a careful look at James' argument reveals that the refutation of SCMF is based on a set of assumptions which yields highly unpalatable consequences. In particular, the fact that James' conclusion is bought at the price of denying the existence of emergent causal unities implies that one has to be very reluctant to use the argument in the context of the present debate between mainstream physicalists and new-wave panexperientialists wherein emergence is taken for granted. So long as the tu quoque argument against panpsychism is based on a naïve endorsement of James' argument, one must conclude that it is no more than a red herring.

## 5 On Aggregates and Unities: Or, What Compounds Might be if not Agglomerative Sums

So far, we have established that GTQP, the backbone of James' assault on MST and on the doctrine of SCMF, yields the absurd conclusion that emergent organized wholes are a chimera. I shall now argue that the reason why James was led to such an absurdity lies in his endorsement of AG, that is, the thesis that, in themselves, all natural "combinations" are no more than mere aggregates. Since genuine unities are more than mere aggregates, it is no surprise that AG necessitates eliminativism about emergent organized wholes. But if such wholes, such unities, exist then the thesis must be false. Yet, if the attempt to constrain organization to linear summation is ill-fated, then there must be metaphysical alternatives to AG and hence there must also be alternative ways in which to formulate the panpsychist hypothesis that advanced forms of consciousness presuppose more primitive forms of consciousness all the way down the cosmological ladder. In the present section I show why the notion of 'aggregate' is antithetical to that of 'unity'; in the next section I conclude with a short discussion of the implications of this finding.

James and Royce were right to maintain that mere aggregates do not constitute intrinsically organized wholes; their mistake, it would seem, lies in the assumption that, in themselves (that is, insofar as their intrinsic properties are considered), all complex natural systems are mere aggregates. It is due to this latter assumption that they were led to the unfortunate conclusion that the only legitimate manner in which a manifold may be said to be "combined" is in the form of an accumulated effect on some tertium quid. Having observed the palpable implausibility of this result, it behooves us to investigate

[18] Surprisingly, such confusion is generic. Thus, it is often the case that the fact (if it is a fact) that the existence of certain features (for example of intrinsic intentionality, or of semantically determined content) is evident in conscious experience is taken as a surrogate for a proper explanation of the theoretical possibility of such features (see Shani (2009a, b) for a criticism of Searle's methodology along such lines).

the aggregativity thesis on which it depends in some more detail. In the previous section I described the unity of organized wholes in terms of their cohesiveness, the latter being an intrinsic property of the macro-organization of a system. To insist that cohesiveness is a real phenomenon is, of course, to insist on the inadequacy of AG. But in order to form a clear impression of just how misguided is the idea that a whole is no more than the agglomerative sum of its proper parts let us take a closer look at the notion of an 'aggregate' and then compare it with cohesive emergent unities as the latter are understood against the background of contemporary complexity theory.

The most conspicuous fact about aggregates is that they are agglomerative sums, linear concatenations of their constitutive elements. Indeed, the affinity between aggregativity and linearity is so vivid that the discussion below begins with two general characteristics of linear functions. Without pretending to be exhaustive, the following list captures a good deal of what it means to be an aggregate[19]:

1) *Order invariance*: The value of linear operations remains invariant amid changes in their order.

To see the point here, consider the notion of a linear function, namely, of a function which satisfies the following properties:

(i) Additivity: $f(x+y) = f(x) + f(y)$.
(ii) Homogeneity (of degree 1): $f(\alpha x) = \alpha f(x)$ for all $\alpha$ (where $\alpha$ is constant).[20]

It is plain that the value of such functions remains invariant amid changes in the order operations (e.g., it matters not whether the summing operation in (i) occurs before, or after the mapping). Also, order invariance follows from the associativity of linear compositions, viz. from the fact that $k \circ (l \circ m) = (k \circ l) \circ m$.

2) *Proportionality*: In linear systems, output is directly proportional to input.

Again, this can be verified by considering a linear function such as $y = 2x + 3$ where the value (output) of the function is directly proportional to the arguments (inputs): an equal rate of change in input always induces an equal rate of change in output, a fact which is reflected in the straight-line geometrical representation of such functions.

3) *Intersubstitution*: Aggregates maintain their identity under rearrangements of their parts.

Mathematically, this condition is satisfied courtesy of the commutativity of linear compositions (e.g., if v and v′ are vectors in a vector space V then $v + v' = v' + v$).

4) *Combinatorial reversibility* (decomposition and re-aggregation): Aggregates can be decomposed and re-aggregated at will without their identity being thereby disrupted.

Mathematically, this condition corresponds to the associativity of linear combinations as well as to the additivity of linear functions. Physically, this means that the system is a-historical, namely, that its evolution is time-independent (symmetric under time

[19] Conditions 3–6 below are adapted, modulo some modifications, from Wimsatt's illuminating discussion of aggregativity (see Wimsatt 1997).
[20] In linear algebra, x and y are vectors whereas $\alpha$ is a scalar.

reversal) and is insensitive to initial conditions. It also means that each part is perfectly separable from any other part; all parts are, as it were, externally related.[21]

5) *Lack of feedback*: The elements of an aggregate exemplify neither cooperative, nor inhibitory interactions. In other words, there are no positive, or negative feedback relations between the parts of agglomerative sums.

Needless to say, this is a clear-cut physical manifestation of linearity.

6) *Qualitative homogeneity*: This condition pertains to aggregative properties. If P (S) is an aggregative property of S, then every part $s_i$ of S has a property $p_i$ corresponding to P, which differs from P only in its magnitude. In other words, aggregative properties are *extensive*, they can be scaled down (or up) recursively while retaining their qualitative character. Moreover, the value of such properties is directly proportional to their amount.

Thus, let P(S) be an extensive property such as volume or (Newtonian) mass. It then follows that the following two conditions are satisfied:

(a) Every proper part $s_i$ of S possesses a property $p_i$ differing from P only in its extent. (b) $P(S) = \sum_{i}^{n} p_i(s_j)$.

As mentioned before, cohesive unities are antithetical to linear aggregates. A brief survey of some of the more significant characteristics of such systems easily reveals the contrast[22]:

To begin, it is a sine qua non of cohesive unities that they are literally constituted by mutually constraining, as well as mutually reinforcing interactions between their constitutive parts, thereby negating condition 5 above. Mutual dependency, whether of the inhibitory or of the synergistic kind is a mark of non-linearity. While cohesive systems may differ widely in their degree of non-linearity, with the behavior of some systems, in some respects, and under certain contexts (for example of certain crystals, or sand piles), approximating that of mere aggregates, it remains true that cohesiveness in all its various manifestations, both in energy-well stabilities and in far from equilibrium systems, always takes the form of a non-localized dynamical balance between interacting elements. In particular, in autonomous systems—from single-cell organisms to socio-culturally embedded persons and to social institutions—cohesiveness is maintained as a delicate, multi-level fabric of non-linear dependencies, that is, of intricate regulatory loops of coordination, inspection, anticipation, task-differentiation, comparison, amplification, restraining, and so on.

From this non-linear mutual dependency many other features of cohesive unities can be shown to be antithetical to those characteristic of aggregates. Consider proportionality (condition 2). This condition is violated for the simple reason that

[21] Internal relations are relations that are essential (viz. necessary) to one or more of the relata. If x is internally related to y, then it cannot be what it is without being so related. For example, the relation between being a husband and having a wife is internal, as is the relation between an angle external to a triangle and the two internal angles non-adjacent to it. By contrast, external relations are inessential for the identity of the relata; each term of the relation can retain its identity in the absence of the relation (e.g. a pebble may be in a box but being in the box is an external relation as both the pebble and the box can retain their identity in its absence). Clearly, combinatorial reversibility dictates that no parts of agglomerative sums are internally related.

[22] The discussion here is indebted to Humphreys (1997), Wimsatt (1997), and in particular Collier (1988, 2008) and Collier and Muller (1998).

some stimuli may cause a chain reaction of mutually reinforcing responses, culminating in a runaway upward spiral (as may happen, for example, in an escalating street riot), while others may be resisted and counterbalanced (e.g., as when a boxer "toughs it up", and ignores the pain in order to survive the round). Indeed, such disproportionality, epitomized in the celebrated butterfly effect, is one of the most recognizable markers of the dynamical evolution of complex systems.

Or, Consider condition six. This condition, which defines aggregative properties, goes against the grain of the very idea of emergent cohesive unities since it denies that which is, arguably, the most conspicuous hallmark of emergent organized wholes, namely, *qualitative novelty*. In the flat world of aggregates and their extensive magnitudes having more (or less) of the same is all one can expect. By contrast, an ontology of organized wholes is characterized by a very different landscape, a landscape with depth, structure, variability, and evolution, where new properties, with novel causal powers, emerge with the advent of new dynamical equilibria of forces and the novel cohesive systems they sustain.[23] As a physically simple example of qualitative novelty brought into existence courtesy of cohesion consider a complex polar molecule (a polar molecule is one where the center of the positive charge does not coincide with the center of negative charge) such as sugar, in which the bonding relations between the constitutive atoms make the molecule an effectively non-invertible dipole, thereby introducing a breakup in the underlying parity symmetry (see Anderson 1972). Another, even more simple example is that of Huygens' discovery that pendulum clocks hung together on a beam become synchronized through mutually transmitted effects and keep better time than either does alone (see Wimsatt 1997). Unlike extensive magnitudes, properties such as polarity and synchronous oscillation are not mere sums of properties of the elements constituting the whole; they are qualitatively distinct from the latter, and their qualitative characters depend on the weaving together of the micro-elements into an organized unity.

Indeed, the denial of qualitative novelty entailed by AG is intimately related to yet another salient feature of aggregates that is antithetical to the nature of cohesive unities, namely, separability.[24] Aggregates are completely atomized; they are nothing but linear sums of totally separate units which remain separate under any possible composition. This theme underlies conditions one, three, and four above, for it is due to separability that aggregates are combinatorially reversible (i.e., de-composable and re-composable), that they remain invariant under inter-substitution of parts, and, more generally, that operations performed on them maintain order invariance.

I will not, in this paper, attempt a systematic critique of the separability thesis since this would require touching upon complex issues ranging from the implications of quantum theory (in particular quantum entanglement, see for example Healy 1991; Humphreys 1997), to the overreliance of Western philosophy on substance metaphysics at the expense of a process-oriented approach (Bickhard 2000a;

[23] An alternative way to describe the difference between a landscape consisting of interacting organized wholes and one populated with aggregates is to say that the former has a form, while the latter consists of a repeatable multiplicity of separate formless elements.

[24] The link, I suspect, is this: qualitative novelty presupposes cohesiveness, hence non-separability, whence separability entails the absence of qualitative novelty.

Campbell 2009), and of modern science on a mechanistic conception of nature (see, e.g., Rosen 1991; Ulanowicz 2000; Whitehead 1925/1997). For our present purpose it is enough to observe the conceptual contrast between this thesis of separability and a description of organized wholes that is in sync with the overtones of contemporary complexity theory. Remarkably, those properties which the separability thesis excludes, such as irreversibility (hence a time-arrow, evolution, and the having of a history), sensitivity to initial conditions, and internal relatedness, all play a very significant role in our current understanding of emergent organized wholes. When these properties are at play, separability becomes a misnomer.

Returning to James, it is interesting to note that separability is deeply ingrained in his critique of SCMF. For, on James' view, each micro-component, be it a material particle, a force, or a mental element, "remains, in the sum, what it always was" (James 1890/1950, 158)—a separable "atom", temporarily associated with other "atoms" in aggregative coalitions which forever fail to qualify as intrinsically organized wholes. And speaking more specifically about mental units, he adds:

> Where the elemental units are supposed to be feelings the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that may mean) still each remains the same feeling it always was, shut in its own skin… (Ibid, 160).

But James is wrong. His contention that parts do not change in the context of the whole is yet another mischaracterization. This can be shown even with regard to one of James' stock examples meant to prove the contrary, namely, water. James argues that 'water' is just the old hydrogen and oxygen atoms put in a new position, without changing a bit in the process, and that the qualitative novelty of water is just their combined effects upon external media (Ibid, 159). But is that so? Consider one such "combined effect", the disposition of water to dissolve salt. Presumably, the capacity of water to break the electrovalent bonds between sodium ions and chlorine ions has to do with the fact that $H_2O$ is a polar covalent molecule, where the electrons making the shared pair are closer to the oxygen, which results in that end of the molecule having a negative charge and the hydrogen end having a positive charge. Now, there are two things that are worth considering here. First, the effect on a tertium quid—disintegrating common salt—is due to polarity, an emergent property of water intrinsic to this chemical compound (neither the hydrogen atoms nor the oxygen atom, considered in isolation, are polarized). Second, the atoms do change in the process—they become (oppositely) charged. True, this change is reversible since one can re-separate the atoms without loosing them in the process, but this does not mean that, while it lasts, the change isn't real (likewise, an individual can be separated from an enflamed crowd and, perchance, calmed down, but this does not change the fact that while in the crowed her behavior and mind-set changed in line with the emergent group dynamics, in ways that were not possible, or likely, otherwise). And, of course, when we shift to the more convoluted, truly organic dynamics of living things—to autopoietic and homeostatic networks of reciprocity—such separability is, in general, no longer available and one is reminded of Aristotle's assertion that the hand without the body is not a hand (*Politics*, I, ii,20).[25]

[25] This is being sourced from B. Jowett (Ed.), *Politics*.

In sum, that aggregates are the complete antithesis of organized wholes is an assertion to which even James would assent. The difference lies in the lesson brought home. For while James believed aggregativity to be the rule and conceded that unity is nowhere to be found save in the eye, or soul, of an observer, the moral of the discussion above is that what needs to be challenged is our tendency to focus on the aggregative, or pseudo-aggregative, aspects of physical reality at the expense of its integrative aspects. After all, as Wimsatt reminds us "[v]ery few system properties are aggregative" (Wimsatt 1997, 382).

## 6 Beyond Agglomeration: Some Afterthoughts About Panpsychism and the Combination Problem

We may agree, then, with James and his present-day followers when they insist that experiences do not sum. But since *no* organized complexes are the direct summands of their constitutive parts we should neither expect macro-experiences to be aggregates of micro-experiences, nor should we loose heart when we discover that they are nothing of the sort. It is remarkable to notice, however, that when critics such as Goff conclude that since there is no intelligible explanation of how micro-experiences aggregate to form macro-experiences it follows that "panpsychism does nothing to explain, in a way that does not appeal to brute emergence, the conscious experience of people and animals" (Goff 2006, 61), they commit a mistake which James himself—who was well aware of the fact that aggregates do not constitute organized wholes as well as of the fact that, unlike MST, there are versions of panpsychism which do not abide by the assumption that macro-experiences are agglomerative sums of micro-experiences—never did commit. The accusation, I have argued, is a veritable red herring.

But while our discussion established that it is wrong to think that James' argument against SCMF constitutes a principled combination problem for panpsychism, the specter of such a problem has not been completely exorcised. As we saw in the opening section, it is possible to think of the combination problem in broad terms such that instead of being identified with the question "how do micro-experiences aggregate into macro-experiences?" it becomes tantamount to the much more general question "does the assumption that even the minutest, most basic physical systems are loci of experience help explain the reality, as well as the character, of sophisticated forms of conscious experience such as those we identify in ourselves and in certain other animals?". When it comes to the explanatory gap, or the hard problem of consciousness, the appeal of panpsychism rests heavily on the possibility of providing a positive answer to this latter question since it is on the promissory note of providing such an answer that panpsychism enjoys some reputations as a vital alternative to standard physicalism. And while the defunct idea that macro-experiences are aggregates of micro-experiences makes the combination problem utterly insoluble the problem (understood in this revised, broad sense) remains a difficult one even if we bid farewell to the aggregativity thesis.

Still, if the explanatory gap is insoluble within a standard physicalist framework (an assumption which is accepted not only by critics of physicalism but also by many of its adherents) then there is a clear motivation for exploring the avenue of

providing a constructive solution to the (generalized) combination problem. Contemporary critics of panpsychism are right, I believe, when they complain that the mere substitution of mild emergence for radical emergence is no guarantee that the emergence of sophisticated forms of experience will then become intelligible. They are, however, too quick when they rush to conclude that such mild emergence is just as brute as the emergence of sentience from an utterly insentient material background (here, again, Goff is a suitable example). The revival of panpsychism in recent years was marked by a genuine and illuminating (if not completely successful) effort to clarify the problematic status of radical ontological emergence. It is time, however, to move beyond this preliminary critical stage and to pay a more careful attention to the question what type of explanation, if any, could possibly account (in an epistemically illuminating fashion) for the evolution of systems with sophisticated experiential life out of the integration of, and interactions between, micro-systems whose degrees of consciousness are asymptotically lower. Only by rising to this challenge would it be possible to determine whether or not the panpsychist alternative can live to the expectation of transcending the epistemic blind alley which seems to plague mainstream physicalism.

The nature of the subject determines that, in the foreseeable future, the problem is bound to remain a speculative metaphysical one rather than metamorphose into a clear-cut scientific puzzle, but this does not mean that it cannot be approached in a systematic and fruitful manner. Panpsychism is an old outlook with deep philosophical roots and these roots afford a richly textured pull of valuable insights, yet such insights must be enriched by the resources available to us at the present, in particular by fresh ideas from quantum physics, complexity theory, and up and coming interdisciplinary research fields such as quantum consciousness.

From a strictly philosophical perspective the most important task, it seems to me, is to try to delineate the general outlines of an *intelligible* transition from simple systems with simple forms of sentience towards more complex systems with more complex forms of experiential qualities. For such a story to be intelligible it is not necessary that the conscious lives of the micro-components of a given system will be *transparent* from the vantage point of the unitary system's conscious perspective, nor is it necessary that zombies will no longer be conceptually possible.[26,27] Rather,

[26] One of the few things which James' argument does show is that *if* a macro-system with a conscious mental life is constituted out of a multitude of micro-systems equipped with their own micro-experiences, and *if* these micro-experiences are instrumental in shaping the experiential character of the macro-system, then these lower-level experiences are *non-transparent* to the conscious perspective of the macro-system. In other words, one cannot access the consciousnesses of, say, one's individual cortical neurons by reflecting on the contents of one's unitary conscious experience. However (and again, pace Goff), I see no reason why transparency should be considered an adequacy criterion in the first place. It is perfectly coherent to assume that an emergent experiential whole is the result of a merging process in which the entity emerging from the merger lacks any privileged access to the mental lives of its constitutive micro-elements. Moreover, for transparency to be possible the constitutive micro-elements must retain their original identity within the whole (hence, to remain separable). As we have seen, there are reasons to suspect that this is often not the case.

[27] As for zombies, the opacity of micro-experiences relative to macro-experiences also explains why it is possible to imagine that a physical duplicate of myself would not have had experiences similar to mine, or even any experiences at all—there simply is no way for us to consciously experience the internal connection between our own experiences and the experiences (if such there be) of, say, our cells and molecules.

the main issue is simply that of providing a plausible account of how systems with primitive capacity for sentience may integrate to form complex cohesive unities whose phenomenal tonality is immeasurably richer.

It is one of the lessons of complexity theory that organization, including in particular the spontaneous proclivity to mold novel organizations, is an intrinsic aspect of nature. If we add to this picture the fundamental panpsychist assumption that nature is also intrinsically sentient then the idea of a self-scaffolding mental "chemistry" suggests itself. Admittedly, by comparison with our knowledge of the outward, material aspects of organization our theoretical knowledge of how inner, experiential property instances may fuse into greater experiential wholes is in its infancy; but to conclude, at this stage, that such integration does not take place, or that it is eternally incomprehensible, is rather premature. Certainly, no argument based on James' critique of MST can establish this pessimistic conclusion. It is hardly surprising that we fail the task of making sense of experiential "self-compounding" so long as we approach the problem using the inappropriate tools of an outdated mechanistic and atomistic view of nature. Rather, in order to put this idea to the test we must first approach it with an open mind, and with our best conceptual tools ready at hand. The jury is still out.

**Acknowledgements** Earlier drafts of this paper were read at the UKZN Spring Colloquium, Byrne Kwazulu-Natal, and at the philosophy department colloquium, University of Johannesburg. I thank the audiences of these meetings for their attendance and helpful comments.

## References

Anderson, P. W. (1972). More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science, 177*, 393–396.

Aristotle. (2009). In B. Jowett (Ed.), *Politics*. LA: Indo-European Publishing.

Barham, J. (2004). The Emergence of Biological Value. In W. A. Dembski & M. Ruse (Eds.), *Debating design: From darwin to DNA* (pp. 210–226). Cambridge: Cambridge University Press.

Bertalanffy, L. V. (1968). *General systems theory*. NY: Braziller.

Bickhard, M. H. (1998). Levels of Representationality. *Journal of Experimental and Theoretical Artificial Intelligence, 10*(2), 179–215.

Bickhard, M. H. (2000a). Emergence. In P. B. Andersen, C. Emmeche, N. O. Finnemann, & P. V. Christiansen (Eds.), *Downward causation* (pp. 322–348). Aarhus: University of Aarhus Press.

Bickhard, M. H. (2000b). Autonomy, function and representation. Special issue on: the contribution of artificial life and the sciences of complexity to the understanding of autonomous systems. *Communication and Cognition—Artificial Intelligence, 17*(3–4), 111–131.

Bickhard, M. H. (2004). The social ontology of persons. In J. I. M. Carpendale & U. Mueller (Eds.), *Social interaction and the development of knowledge* (pp. 111–132). Mahwah: Erlbaum.

Bickhard, M. H. (2006). Developmental Normativity and Normative Development. In L. Smith & J. Voneche (Eds.), *Norms in human development* (pp. 57–76). Cambridge: Cambridge University Press.

Bickhard, M. H., & Campbell, D. T. (2003). Variations in variation and selection: The ubiquity of the variation-and-selective retention ratchet in emergent organizational complexity. *Foundations of Science, 8*(3), 215–282.

Campbell, R. (2009). A Process-based Model for an Interactive Ontology. *Synthese*, Special Issue on Interactivism (Ed. M. Bickhard), 166, 3, 453–477.

Capra, F. (1996). *The web of life: A new scientific understanding of living systems*. NY: Anchor Books.

Carruthers, P., & Schechter, E. (2006). Can panpsychism bridge the explanatory gap? *Journal of Consciousness Studies, 13*(10–11), 32–39.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.
Chalmers, D. J. (2002). Consciousness and its Place in Nature. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 247–271). NY: Oxford University Press.
Christensen, W. D., & Hooker, C. A. (2000). An Interactivist-constructivist approach to intelligence: Self-directed anticipative learning. *Philosophical Psychology, 13*(1), 5–45.
Clifford, W. K. (1878). On the nature of things-in-themselves. *Mind, 3*(9), 57–67.
Collier, J. D. (1988). Supervenience and reduction in biological hierarchies; Philosophy and Biology. In: M. Matthen and B. Linsky (eds.) *Canadian Journal of Philosophy* Supplementary Vol. 14.
Collier, J. D. (2008). A dynamical account of emergence. *Cybernetics and Human Knowing, 15*(3–4), 75–86.
Collier, J. D., & Muller, S. J. (1998). The Dynamical Basis of Emergence. In G. Farre & T. Oksala (Eds.), *Emergence, complexity, hierarchy and organization, selected and edited papers from the ECHO III conference, acta polytechnica scandinavica, MA91*. Espoo: Finish Academy of Technology.
de Quincy, C. (2002). *Radical nature: Rediscovering the soul of matter*. Montpelier: Invisible Cities Press.
Goff, P. (2006). Experiences don't Sum. *Journal of Consciousness Studies, 13*(10–11), 53–61.
Griffin, D. R. (1998). *Unsnarling the world knot: Consciousness, freedom, and the mind-body problem*. Berkeley: University of California Press.
Healy, R. A. (1991). Holism and Nonseparability. *Journal of Philosophy, 88*(8), 393–421.
Henle, M. (1990). William james and gestalt psychology. In M. G. Johnson & T. B. Henley (Eds.), *Reflections on the principles of psychology*. Hillsdale: Laurence Elrbaum Associates.
Humphreys, P. (1997). How properties emerge. *Philosophy of Science, 64*(1), 1–17.
James, W. (1890/1950). *The Principles of Psychology, Vols. 1 & 2*. NY: Dover Publications.
Jantsch, E. (1976). Evolution: Self-realization through self-transcendence. In E. Jantsch & C. H. Waddington (Eds.), *Evolution and consciousness: Human systems in transition*. Reading: Addison-Wesley.
Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. NY: Oxford University Press.
Kauffman, S. A. (2000). *Investigations*. New York: Oxford University Press.
Laszlo, E. (1996). *The systems view of the world: A holistic vision for our time*. Cresskill: Hampton Press.
Nicolis, G., & Prigogine, I. (1989). *Exploring complexity: An introduction*. NY: W.H. Freeman.
Prigogine, E. (1978). Time, structure, and fluctuations. *Science, 201*(4358), 777–785.
Prigogine, E., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam.
Rosen, R. (1991). *Life itself: A comprehensive Inquiry into the nature, origin, and fabrication of life*. NY: Columbia University Press.
Routh, J. I. (1965). *Fundamentals of inorganic organic and biological chemistry*. Philadelphia: W.B. Saunders.
Royce, J. (1881). "Mind-Stuff" and Reality. *Mind, 6*(23), 365–377.
Ruiz Mirazo, K., & Moreno, A. (2000). Searching for the roots of autonomy: The natural and artificial paradigms revisited. *In CC AI: Communication and Cognition-Artificial Intelligence, 17*(3–4), 209–228.
Russell, B. (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.
Seager, W. E. (1995). Consciousness, information, and panpsychism. *Journal of Consciousness Studies, 2*, 272–288.
Shani, I. (2006). Narcissistic sensations and intentional directedness. *Cybernetics and Human Knowing, 13*(3), 87–110.
Shani, I. (2007). The myth of reductive extensionalism. *Axiomathes, 17*(2), 155–183.
Shani, I. (2009a). The whole rabbit: On the perceptual roots of Quine's indeterminacy puzzle. *Philosophical Psychology, 22*(6), 739–763.
Shani, I. (2009b). Representation and aspectual shape. Perspectives on interactivism, a special issue of *New Ideas in Psychology* (Ed. A. Levine). doi:10.1016/j.newideapsych.2009.09.007.
Skrbina, D. (2006). Beyond descartes: Panpsychism revisited. *Axiomathes, 16*, 387–423.
Steiner, P., & Stewart, J. (2009). From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and the Cognitive Sciences, 8*, 527–550.
Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research, 62*, 253–281.
Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies, 13*(10–11), 3–31.

Ulanowicz, R. E. (2000). Life after newton: An ecological metaphysics. In D. R. Keller & F. B. Golley (Eds.), *The philosophy of ecology*. Athens: University of Georgia Press.
Van Cleve, J. (1990). Mind-Dust or Magic? Panpsychism versus Emergence. *Philosophical Perspectives*, vol. 4, Action Theory and Philosophy of Mind, 215–226.
Várela, F. J. (1979). *Principles of biological autonomy*. New York: Elsevier.
Whitehead, A. N. (1925/1997). *Science and the modern world*. NY: The Free Press.
Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64*, 372–384.