

WILEY

International Phenomenological Society Philosophy and Phenomenological Research

Mind/Consciousness Dualism in Sāṅkhya-Yoga Philosophy

Author(s): Paul Schweizer

Source: *Philosophy and Phenomenological Research*, Vol. 53, No. 4 (Dec., 1993), pp. 845-859

Published by: [International Phenomenological Society](#)

Stable URL: <http://www.jstor.org/stable/2108256>

Accessed: 02-12-2015 16:31 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley, International Phenomenological Society and *Philosophy and Phenomenological Research* are collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

<http://www.jstor.org>

Mind/Consciousness Dualism in Sāṅkhya-Yoga Philosophy*

PAUL SCHWEIZER
University of Edinburgh

1 Introduction

In the western philosophical tradition, there is a very prevalent tendency to equate the mind with the conscious self. This tendency is exemplified in the framework of Descartes' *Meditations*, which, notwithstanding the abundant criticism this text has received over the centuries, in some sense still provides the most definitive and well articulated expression of the presuppositions underlying the modern western conception of mind. Descartes unhesitatingly identifies the referent of the indexical expression 'I' with *res cogitans*, the thinking substance, and thereby conceives the mind and self as one. For Descartes, thinking substance is of course metaphysically independent of extended substance, which engenders his familiar (and surprisingly resilient) dualism between mind and matter. Also without hesitation, Descartes takes conscious thought and subjective experience to be unqualified constituents of mental substance, and this particular mode of classification has become so much a part of our conceptual heritage that the dividing line he draws (even if not the attendant metaphysical commitment) seems almost self-evident. Surely, if a dualism with matter is to be entertained, then thoughts and subjective presentations belong on the *non-material* side of the dividing line.

In contrast to this fairly ubiquitous Cartesian view, I would like to examine an alternative but intimately related version of substance dualism advanced by two allied schools (*darśana*) of orthodox Hindu philosophy, namely the Sāṅkhya school, and classical Yoga as expounded by Patañjali.¹

* I would like to thank an anonymous referee of *PPR* for helpful comments.

¹ By using the appellation 'Sāṅkhya-Yoga' philosophy, I do not mean to endorse the view that the two schools are essentially the same, nor that Yoga is simply an applied or practical component grafted onto Sāṅkhya theory. Rather, I merely wish to discuss some (of the many) theoretical points which the two schools have in common, though I have tended to use predominantly Sāṅkhya terminology.

These schools draw a metaphysical dividing line not between mind and matter, but rather between matter and consciousness, wherein the mind is placed on the *material* side of the ontological divide. I think that this alternate formulation of the mind/body problem constitutes a valuable intellectual perspective for contemporary western philosophy. It illuminates the problem from a different angle, and this in turn highlights some of the presuppositions underlying the western analysis, and reveals that in an important sense these presuppositions constitute *choices* about how to conceptualize the phenomena, rather than inherent divisions supported by the phenomena themselves. In particular, the Cartesian identification of mind with the conscious subject will be brought into sharp focus, as well as the related Humean identification of the self with a constellation of subjective ideas.

In section 2 of the present paper, I will give a general overview of the Indian position, and in sections 3 and 4 I will explore some of the theoretical consequences of this alternative dualism, and critically compare these consequences with some recent manifestations of the Cartesian model. A primary tenet of the paper is that the classical Indian division between pure consciousness and matter has distinct philosophical interest, especially with respect to current discussion on the relationship between consciousness and mental representation, as well as the endeavor to give a computational and/or naturalistic account of mentality.

The philosophical interest of this alternative dualism is at least partly due to the fact that Descartes' schism between mind and matter has proved extremely adaptable, and reappears under a variety of contemporary guises. For example, Searle (1980) accuses proponents of 'strong' AI (artificial intelligence) of embracing a rigid separation between mind and matter, since they view the mind as a program of rules for symbol manipulation, and this is to treat the mind as an abstract formalism, fundamentally independent of the physical medium in which it happens to be instantiated. Thus certain advocates of the very recent discipline of AI seem committed to a strict separation between the mental and the physical. In a related general vein, the computational and formal treatments of mind advocated by cognitive science and AI are by nature syntactical, even when considered at the level of physical instantiation, and as such they seem to say nothing about the 'real meanings' or semantical contents which are quintessentially mental. Indeed, the deep theoretical gulf between syntax and semantics can be viewed as a highly specialized and refined variation on Descartes' theme. Syntax (at the level of actual tokens) is a system of physical characters, while semantical content seems to be something distinctively *cognitive*, and the connection between the two, at least with respect to computational theories of mind and natural language processing, is a very problematic issue.

Another well known guise in which standard Cartesian dualism finds expression is the infamous gap between qualia, *i.e.*, the ‘elements’ of subjective conscious experience, and the associated neurophysiological configurations of the brain. Subjective ‘ideas’ or presentations are held by Descartes to belong to the realm of *res cogitans*, whereas the brain obviously belongs to the realm of *res extensa*. This gap between qualia and brain states constitutes a theoretical focal point of modern dualism, since the correlation between objective physical states of the brain and subjective conscious experience is the most intimate level of correspondence between matter and mind, and it seems to form the crucial boundary line where the two substances should ‘intersect.’ As I will argue below, the Sāṅkhya-Yoga analysis, though still dualistic, is less problematic with regard to the representational content of conscious mental states, as well as with regard to the functional role of qualia, since content and consciousness are held to be metaphysically independent.

2 Sāṅkhya-Yoga Dualism

The Sāṅkhya school, or *darśana*, is one of the oldest philosophical traditions of India, and many of its ideas are traceable to the Ṛgveda and the early Upaniṣads. Its historical founder is Kapila, though the original *Sāṅkhyasūtras* (aphorisms) he is said to have written during the 6th or 7th century B.C. are now lost, and the most important of the existing texts is the *Sāṅkhyakārikā* (explanatory verses) of Īśvarakṛṣṇa, from around the 3rd century A.D. The Sāṅkhya tradition has a great many theoretical points in common with the classical Yoga *darśana* as expounded in Patañjali’s *Yogasūtra*, probably written somewhere from the 4th to the 2nd century B.C.,² and the metaphysical position discussed in the present paper is part of their shared philosophical framework. Thus the basic dualism between consciousness and matter, as well as more specific allied points, will henceforth be referred to, somewhat generically, as the ‘Sāṅkhya-Yoga view.’

According to the Sāṅkhya-Yoga view, the ultimate principle underlying matter is *prakṛti*, the metaphysical substrate supporting all material phenomena.³ The mind is deemed to be part of the material world, and hence to be metaphysically grounded in *prakṛti*. The body is also part of the realm of *prakṛti*, and thus mind and body are held to be of the same metaphysical

² Textual chronology is of course a notoriously difficult and controversial affair. The dates provided in this paragraph are taken from Raju (1985).

³ In the ensuing discussion, the terms ‘*prakṛti*’ and ‘matter’ will be used roughly as synonyms, even though this is not strictly correct, if matter is construed in terms of the ‘particles’ which make up physical objects. *Prakṛti* is the *metaphysical* principle which underlies physical manifestations, though for expository convenience I will often equate the physical world with *prakṛti*.

substance. In its most general connotation, the term 'mind' in western philosophy corresponds to the combination or complex of three faculties recognized in the Sāṅkhya-Yoga taxonomy, namely *manas*, *buddhi*, and *aḥamkāra*, whose respective significance will be briefly outlined below.

Manas (which is often translated directly as 'mind', though it is only a single facet of the 'mental triplex') is viewed essentially as an organ, the special organ of cognition, just as the eyes are the special organs of sight. Indeed, *manas* is held to be intimately connected with perception, since the raw data supplied by the senses must be ordered and categorized with respect to a conceptual scheme before various objects can be perceived as members of their respective categories, and as inhabiting a world characterized by the systematic and distinguishable attributes normally perceived. This imposition of conceptual structure on the chaotic field of raw sensation is one of the basal activities of *manas*, and forms the distinction between brute sensation (*nirvikalpaka*) as opposed to differentiated perception (*savikalpaka*). Hence ordinary perceptual experience is already heavily conditioned by the activities of *manas*, and *manas* is thus sometimes referred to as the sixth organ of sensation.

In addition to its perceptual activities, *manas* is held to be responsible for the cognitive functions of analysis, deliberation and decision. It is closely allied to *buddhi*, which is somewhat roughly translated as the faculty of 'intellect' or 'reason.' *Buddhi* is a subtler and more powerful faculty than *manas*, and is responsible for the higher level intellectual functions, which require intuition, insight and reflection. The Indian *buddhi* is in some ways comparable to the Greek *nous*, while *manas* is responsible for lower level discursive thought and analysis. But *buddhi* is still regarded as a manifestation of *prakṛti*, albeit the most subtle and refined form which material substance can assume. The combination of *manas* and *buddhi* roughly correspond to what is meant by the objective or 'impersonal' mental faculties in western philosophical discourse. In addition, Sāṅkhya-Yoga recognizes a third component of mind, *aḥamkāra*, which is the ego or phenomenal self. *Aḥamkāra* appropriates all mental experiences to itself, and thus 'personalizes' the objective activities of *manas* and *buddhi* by assuming possession of them. The combination of these three faculties is referred to as *antahkaraṇa*, the 'inner instrument,' which approximately comprises the individual mind-self of the western philosophical tradition.

But on the Sāṅkhya-Yoga account, the realm of *prakṛti* or matter is held to be inherently unconscious, and is thereby incapable of producing consciousness as an effect. The manifestations of *prakṛti* are always objects, and it is argued that objects can never transform themselves into subjects. Thus at the heart of this dualistic position is the notion that mind-material is not capable of *generating* consciousness out of unconscious ingredients.

Subjective awareness is a distinct ontological category, and in principle it cannot be derived from the stuff of which objects are made. So, in sharp contrast to the western approach, the mind and the cognitive activities it sustains are held to be intrinsically unconscious, since *manas*, *buddhi* and *aḥamkāra* are all manifestations of *prakṛti*.

According to the ancient view, the dynamics of *prakṛti* are governed by the interactions of the three *guṇas*, which are the three basic types of constituent of physical substance. The three *guṇas* are *sattva*, *rajas* and *tamas*, which correspond roughly with 'transparency and buoyancy,' 'energy and activity,' and 'inertia and obstruction.' All physical phenomena are believed to consist of unstable mixtures of these three types of constituent, and the instability of these mixtures is responsible for the evolution and transformations of the material world. Thus the conceptual processes sustained by the mind are governed by the mechanical and unconscious interplay of the *guṇas*, and to this extent, mental phenomena are viewed in purely 'physicalistic' or mechanical terms. The unfolding of thought-forms is an integral part of the evolution of *prakṛti*, and mental processes are simply the result of appropriate transformations of unconscious material substance. It is worth noting at this point that the Sāṅkhya-Yoga view thereby avoids one of the most serious pitfalls of Cartesian dualism, since on the Indian account, mental causation does not violate physical conservation laws. By including the mind in the realm of matter, mental events are granted causal efficacy, and are therefore able to directly initiate bodily motions.

Consciousness, on the other hand, is held to belong to a different ontological category altogether. Consciousness is placed in the realm of *puruṣa*, the absolute, unconditioned self, which in some respects is comparable to Kant's noumenal self. *Puruṣa* is described as pure and undifferentiated awareness, and it is held to be immutable and inactive, to be formless and without parts or limiting characteristics. *Puruṣa* is the metaphysical principle underlying the individual person, and closely corresponds to the *ātman* of the Vedānta school.⁴ *Puruṣa* is held to exist in complete independence of the material realm, and so the basic dualism in the Sāṅkhya-Yoga metaphysics is between *puruṣa* and *prakṛti*, between consciousness and matter. On this account, mind belongs to the world of matter, while the self belongs to the realm of pure awareness, and thus the mind and the self are ontologically distinct. Movement and form are characteristics of matter, and they are also characteristics of thought, which is a manifestation of matter, while consciousness is held to be intrinsically formless and unchanging.

⁴ It should be noted, however, that Sāṅkhya-Yoga recognizes a multiplicity of distinct *puruṣas*, i.e. one for each self, while Vedāntic thought tends to view the *ātman* as ultimately singular. Hence on the former account, there is a numerically distinct *puruṣa* associated with each material mind-complex.

What then is the relationship between consciousness and matter, and how are conscious mental events possible? According to the Sāṅkhya-Yoga model, thought processes and mental events are conscious only to the extent that they receive external 'illumination' from *puruṣa*. Consciousness is standardly compared to a light, which illuminates the specific material configurations or 'shapes' assumed by the mind. It is the subtle 'thought-material' of the *buddhi* which allows mental events to appear conscious, because the refined *buddhi* substance is transparent to the light of consciousness. Thus conscious thoughts and perceptual experiences take place when *buddhi* receives representational forms, both perceptual and conceptual, from *manas*, the organ of cognition. *Buddhi* is believed to consist of a preponderance of the 'transparent' *sattva guṇa*, and thus the representational forms it receives from *manas* are capable of becoming translucent with the light of consciousness. So *buddhi* receives cognitive structures from *manas*, and conscious 'light' from *puruṣa*, and in this manner, specific mental structures are capable of being illuminated by an external source, and thereby these structures are able to appear conscious. But consciousness itself is entirely independent of the particular thought structures it happens to illuminate.

The translucent quality of *buddhi* distinguishes thought-stuff from the gross material objects of thought and perception, which ordinarily contain a preponderance of the 'dark' or opaque *tamas guṇa*. Thus only the thought-material of *buddhi* is capable of conscious illumination, rather than the entire material realm, since a preponderance of the *tamas guṇa* renders the external objects of the material realm opaque to the light of consciousness. Hence representational structures, rather than the objects represented, are capable of conscious illumination, which is why minds appear sentient while stones and tables do not. Only the subtle stuff of the mind is a suitable medium for receiving sentience, and thus it is minds which are the loci of awareness in the natural world. In this manner, conscious thought processes and attendant representational content is comparable to photographic slides of external objects. The photographic image stored in the film is composed of matter, but it is both representational and translucent, and therefore is analogous to the thought-structures which glow with the sentience of *puruṣa*.

So, on this model, the representational content of thought is carried in the unconscious physical configurations of the mind, and certain of these patterns or configurations become illuminated by an external and undifferentiated awareness, resulting in the phenomenon of particular conscious thoughts and subjective experiences. But consciousness and the illuminated mental processes are entirely independent. Formless and immutable consciousness plays no causal role in the transformation of mental structures, but rather is a passive 'witness' to some small portion of these structural

transformations, viz. that portion normally identified with the conscious individual or Cartesian ego. According to the Sāṅkhya-Yoga analysis, the Cartesian or phenomenal ego is not the real self, but rather is a partial reflection of the self in the determinate structures of the material *buddhi*. The real self, as pure awareness, cannot be experienced as an object of consciousness, since all direct objects of consciousness are held to be specific structures of *sattvic prakṛti* illuminated by *puruṣa*. Thus the constellation of subjective experience which comprises Hume's 'empirical' self is held to be a mere material pattern, which receives 'transcendental' illumination from *puruṣa*.

Indeed, the habitual identification of the individual subject with the illuminated transformations of *buddhi* is traditionally diagnosed as a particular mode of *buddhi*, and is held to be the root form of *avidyā* or ignorance regarding one's true nature. A standard illustration of the purported relation between the real self and the ego is the following. The sun, as reflected on the surface of the ocean, appears to fluctuate with the waves and agitations of the surface, but it is only the reflection, rather than the actual sun, which is affected by these disturbances. Thus the phenomenal self undergoes innumerable changes, and possesses the varied structural attributes of its physical medium, but these are true only of the reflection of self in an external substrate. According to the Sāṅkhya-Yoga account, the real self is not to be identified with the Cartesian realm of qualitative experience, and there simply is no *res cogitans*, no thinking substance in Descartes' sense. It is material substance which thinks, and it is the self which makes these thoughts conscious.

3 The Distinction Between- Consciousness and Content

The foregoing brief sketch of the Sāṅkhya-Yoga view presents a striking contrast, on a number of points, with the traditional western analysis of self and mind. For my present purpose, the most salient point of contrast concerns the theoretical separation between consciousness and mental representation, rather than questions concerning the proper analysis of the self. In the ensuing discussion, I would like to examine some consequences of a general mind/consciousness distinction, as inspired by the ancient Indian *darśana*, but I would like to examine these consequences from the more contemporary perspective of mental representation in cognitive science and AI. Thus not all of the following claims and considerations are held to be contained in, nor to be strictly deducible from, the traditional Sāṅkhya-Yoga philosophy; rather I wish to explore some features of interest which follow from a broad conceptual alternative to the Cartesian picture, the latter of which still colors much contemporary thought on these subjects.

On the Cartesian view, subjective presentations or 'ideas', as *structured objects*, are held to belong to the non-material realm. But, at the same time, they are held to represent external material objects and states of affairs. This leads to an immediate and very serious difficulty, since if the two spheres are metaphysically independent, then it is highly unclear how events in the mental realm could be related to or systematically represent objects in the material realm. Thought becomes separated, by an ontological chasm, from the external objects of thought, and the inner world of subjective experience threatens to become a solipsistic bubble. If some sort of (divinely harmonious) correlation between the two spheres is advocated, in order to provide a foundation for epistemology, then this leads to a profligate and un-aesthetic version of metaphysical repetition, where objects in *res extensa* are gratuitously mimicked by their representational duplicates in *res cogitans*.

Mind/consciousness dualism, though obviously not free from difficulties, at least possesses a theoretical advantage over traditional western dualism in this respect, since the conscious aspect of subjective experience is entirely disengaged from its semantical or representational form. The epistemological and cognitive advantages of this separation are most clearly illustrated in the 'base level' cases of veridical perception, and singular thoughts involving objects encountered in the environment. In these cases, representational structures are of the same metaphysical category as the objects they represent, just as a chair and a photograph of a chair are both composed of matter. This metaphysical (and hence causal) homogeneity provides a theoretical foothold for naturalistic accounts of representational content, since, just as the photograph of a chair can be accounted for *via* the assorted optical and chemical processes involved in its formation, so the representational content or structure of the perception of the chair can (in principle at least) be accounted for in terms of the physical interactions and processes involved in *its* formation.

It is the conscious, subjective aspect of visual perception which serves to motivate the introduction of a distinct metaphysical category, not the causally induced representational structure of perception, since it is theoretically feasible that the latter can be explained in terms of *unconscious* mechanisms, of generally the same sort that would be applied in the case of robotic 'vision.' The deep philosophical problem in the case of human perception lies not in the explication of structure, but rather in the fact that this bio-mechanically induced structure is imbued with conscious awareness. Thus Descartes' dualism of mind and matter, wherein conscious presentations *qua* structured objects are placed in the purely non-material realm, embodies an unhappy conflation of theoretically distinct features, and it thereby creates a causal abyss which insulates perceptual experience from the material objects which are perceived. It is consciousness, rather than content, which provides the most compelling impetus for dualism.

Thus the Sāṅkhya-Yoga division is based on a more perspicuous distillation of the salient components of subjective experience, where the world of qualia, the 'Cartesian theatre' of conscious presentation, is analysed as a mixture rather than as a realm of pure substance. The representational content of qualia is due to the respective material structures or patterns which comprise them, while conscious illumination is external, and ontologically independent of structure. Conscious illumination of structure is made possible by the particular *composition* of the pattern, wherein the transparent material of thought-stuff enables the light of *puruṣa* to enter. Thus representational content is comparable to the geometrical patterns of a stained glass window, which exist independently of their illumination by the sun. The illuminated patterns are a composite, consisting both of structured glass, which is analogous to the thought-shapes assumed by *sattvic buddhi*, and of luminescence, which is the consciousness of *puruṣa*. In this manner, qualia are seen as composite phenomena, and, in the case of perceptual experience, there is no ontological chasm separating representational objects from the objects represented. Instead, it is the underlying conscious illumination, common to all experience, which defines the metaphysical boundary line between subject and object.

It is important to note that the foregoing critique of the Cartesian picture is not based on a general philosophical tenet to the effect that a metaphysical homogeneity *must* obtain between a representational object and the object represented, and which would entail that the mind must be material if it is capable of representing material entities. This inversion of Berkeley's argument against matter is far too stringent, and indeed would face serious difficulties in the case of abstract and non-specific objects, universals, non-existents, etc. Rather my point is simply that representational content, even for a cognitive phenomenon as basic as visual perception, is not susceptible to naturalistic explanation, if this content is held to reside in a non-material medium. This point is intimately linked to the earlier observation concerning mental causation with respect to bodily motions, since the problem of interaction between disparate substances cuts both ways. However, if mind and environment are held to belong to the same metaphysical realm, then mental content can both cause and be caused by other physical events. This at least opens the door to explaining mental representation and the evolution of cognitive structure through appeal to the interaction between an organism and its environment, while it is not at all clear that this door is open on a Cartesian account.

These considerations suggest that the ancient Indian separation between conscious awareness and representational content has interesting ties to current issues in cognitive science and philosophy of mind. As will be discussed below, a crucial relation between human and computer-based intelligence is

brought into sharp focus by mind/consciousness dualism, while this same relation is obscured by the familiar Cartesian divide. The research programmes of cognitive science and AI are based on a computational paradigm, in which it is assumed that cognitive phenomena, both natural and artificial, are founded on computational procedures instantiated in physical systems. The material view of mind endorsed by the Sāṅkhya-Yoga analysis is in principle quite compatible with this paradigm, since it is conceivable that the cognitive activities of *manas-buddhi* can best be described in computational terms. *Manas-buddhi* is an unconscious mechanism which manipulates the various representational structures involved in perception, cognition, and language, and from a late 20th century perspective, these manipulations can perhaps most plausibly be characterized as computational.

In this respect, there is a very strong resemblance between the activities of the cognitive organ of mind, and the syntactic manipulations carried out by a computer. Given a sufficiently sophisticated robot, *viz.* one which could perform 'autonomously' in a rich and complex behavioral environment, and thereby pass some suitably rigorous version of the Turing test, there would be many theoretical considerations in favor of attributing mentality to such an artifact. From an abstract perspective, there does not seem to be a significant difference between the mechanical activities of *manas-buddhi* and the computational procedures of an 'artificially intelligent' system. This indicates that the Sāṅkhya-Yoga position enjoys a very close fit with modern functionalist accounts of mind.

The fit is especially close in that the Sāṅkhya-Yoga analysis of mind makes no allusion to the presence of conscious experience. I think that, conversely, it is quite clear that functionalism and the computational paradigm have nothing important to say about consciousness, and that subjective experience is an element which is theoretically extraneous to the research programmes of cognitive science and AI.⁵ Computational procedures are mathematical abstractions, defined over purely formal objects, without reference to consciousness, and there is little reason to believe that consciousness or genuine subjectivity will simply 'emerge' from instantiated procedures of sufficient complexity or subtlety. Indeed, it is only the representational structure of qualitative experience which is relevant to its functional role, and hence this role can always be filled by an object with the same functional structure but which by hypothesis is *not* conscious. Since

⁵ Various 'missing qualia' arguments, such as Block (1978) and Jackson (1982), have objected that functionalism is not in principle able to account for subjective experience, and I would agree with the view that considerations of functional structure alone cannot provide a sufficient condition for consciousness. On the dualistic picture expounded in the present work, functional structure is not even a sufficient condition for the ability of an instantiated system to *receive* consciousness, since receptivity is a feature specific to composition.

there is no contradiction inherent in this last claim (unless one is smuggled in by begging the question), it follows that consciousness *per se* is theoretically superfluous to functionalism.

4 Cartesianism and the Chinese Room

Searle (1980) provides a well known and important criticism of the classical AI position, a position which asserts that computational models have explanatory significance with respect to genuine mental states. Searle employs his celebrated Chinese room thought experiment in an attempt to demonstrate that instantiating an abstract system of rule-governed symbol manipulation is *not* a sufficient condition for the attribution of mentality. In brief, Searle argues that programs are merely recipes for grinding syntax, while genuine understanding requires an intrinsically semantical or intentional component. Thus no system can possess genuine understanding simply in virtue of instantiating a program, and therefore algorithmic processes cannot provide an adequate account of the mind. He insists that *human* understanding is inextricably tied to the particular physical 'stuff' out of which brains are made, and thus functional structure alone cannot account for intentionality, since functional structure is by definition independent of its medium of instantiation.

In the context of the present discussion, it is interesting to look a bit closer at Searle's claim that the physical stuff of which brains are made is crucial to their ability to sustain the intentional aspect which underlies true understanding. In (1990), Searle further clarifies his position by explicitly equating intentional content with the ability to be consciously entertained by a subject. This equation sheds a very revealing light on the Chinese room scenario, because it shows that what is most obviously missing from the Chinese room is *conscious presentation of representational content*, which Searle in turn takes to be essential for genuine understanding. I think that the intuitive force behind Searle's objection to the doctrine of strong AI lies in the fact that the Chinese room lacks conscious presentations, and specifically, it lacks presentations of the representational content associated with the system's ability to pass the Chinese Turing test. Searle, as homunculus in the room, only has presentations regarding his understanding of English, and of his attendant ability to execute the program, written in English, which tells him how to manipulate the 'meaningless' Chinese symbols. This lack of 'semantical presentation' with respect to the Chinese symbol processing constitutes the real asymmetry between English and Chinese, and thus the pivotal issue in the thought experiment is conscious presentation, which we may or may not wish to equate with 'real understanding.'

So, according to this reading of Searle's argument, lack of conscious presentation can be isolated as the crucial element missing from the Chinese

room, and failure to reproduce the specific material composition of the brain can be isolated as the *cause* of this lack. Hence potential isomorphism of functional structure has been granted, while genuine mentality (*i.e.*, presence of associated conscious presentations) has been denied, on grounds which ultimately reduce to the specific material composition of the brain. This gives rise, on the one hand, to a striking similarity between the Searlean critique of strong AI and the Sāṅkhya-Yoga analysis of mind, although it also highlights a crucial difference between the two, and reveals a strong neo-Cartesian element in Searle's position. The resemblance will be discussed first.

As described in the previous section, it is a basic tenet of the Sāṅkhya-Yoga view that the specific composition of the mind's thought-structures is essential to whether they can receive conscious illumination. It is only the very special *sattvic* characteristics of *buddhi* which allow human thoughts and experiences to 'absorb' consciousness. Thus, according to a functionalist rendition of Sāṅkhya-Yoga dualism, the actual medium of instantiation is essential to whether representational content can support conscious awareness. It is the particular compositional details of human, brain-supported thought-stuff which make awareness of content possible, and to this extent, Sāṅkhya-Yoga dualism is in accord with Searle in maintaining that isomorphism of functional or representational structure is not a sufficient condition for reproducing subjective experience, and thus that the Chinese room cannot possess all the characteristics essential to the human mind-self complex. Functionalist methodology is capable of reproducing the mechanical characteristics of mind, which, when suitably instantiated, would allow an artificial system to completely simulate all behavioral aspects of human beings. If the mind is viewed as computational, while consciousness is held to be independent of computational structure, then it is possible to maintain simultaneously that the research paradigm of cognitive science should be capable of accounting for all phenomena which are properly cognitive or mental, *and* that functionalism does not provide a sufficient condition for the presence of qualitative experience.

In contrast to the Searlean view, mind/consciousness dualism would deny that conscious presentation has any intrinsic link to representational content. Representational content is defined in purely functional terms, and those terms, once instantiated, are sufficient to account for the causal and behavioral interactions of all 'intelligent' physical systems, including human beings. In this manner, there is no properly *cognitive* difference between human qualia, and functionally isomorphic structures exploited by unconscious machines. If mind and consciousness are separated and the mind is construed computationally, then there can be no appeal to a uniquely 'semantical' component in the characterization of genuine understanding.

Computational or formal procedures are by nature syntactical, even though the structures and manipulations involved may be *interpreted* as playing representational or semantical roles. For example, biological or design based heuristics may, for various reasons, attribute representational content to the structures over which computations are performed, where these reasons will advert to patterns of actual or potential interaction between the system and its environment. This is an operational account of content which does not correspond to the sense of 'intrinsic aboutness' which Searle attempts to capture through his appeal to conscious presentation. However, on the model of mind/consciousness dualism, conscious awareness can not in itself imbue a mechanical system with 'intrinsic aboutness.' The fact that some human representational structures are consciously presented does not give them the 'magical' power to uniquely determine their (operationally specified) objects in the external world. Hence consciousness does not confer a privileged semantical dimension to the mind, but, on this account, merely illuminates some of the syntax.

So it would appear that any appeal to conscious presentation, in order to explicate the type of 'real' mental content which purportedly distinguishes the human mind from functionally equivalent artifacts, constitutes a tacit invocation of *res cogitans*. Of course, most contemporary proponents of this Cartesian heritage would not be willing to posit a distinct ontological realm in which subjective ideas reside, but rather would hold the familiar components of mental substance to be somehow exuded by relevant organic processes, thereby endorsing a 'naturalized' version of the Cartesian picture. Yet in many ways this constitutes an unfortunate move, since the severe problems which afflict Descartes' notion of mental content indicate that this notion is better abandoned rather than naturalized. For example, if genuine semantical or representational content is equated with the potential to be consciously entertained, then the deeply troublesome issue concerning privileged introspective access to the mental realm is re-introduced, and in fact assumes central theoretical importance, since potential awareness then becomes an essential criterion for meaning. This problem is especially acute in an attempted naturalistic setting, since it is not clear that the introspective awareness of meaning has *anything* to do with the causal processes responsible for physical behavior, or with the many irredeemably *unconscious* processing states of the brain which intercede between the various episodes of conscious thought.

Even on the physicalist hypothesis that consciousness is nothing but a product of material brain state, it still does not follow that these particular brain states are of unique causal or functional significance. Still less does it follow that they are causally significant *by virtue of* the physical properties responsible for consciousness. Thus there does not appear to be a naturalistic or causal motivation for the identification of 'genuine' representational con-

tent with the Cartesian realm of conscious presentation. And this leaves open the possibility that, even on a physicalist account of awareness, the semantical content which is presented has little or nothing to do with the explanatorily significant brain states which govern both the physical behavior of the system and the continued evolution of the internal sequence of brain states. Indeed, within a naturalistic setting, it is difficult to see why the presumed self-transparency of *res cogitans* should have any causal significance, even with respect to the transition from one conscious state to the next.

Mind/consciousness dualism would of course deny that consciousness has privileged introspective access to content (since content is independent of awareness) and it would also deny that consciously presented content, as such, has any unique semantical or causal status. A central feature of this model is that there is no intrinsic representational or functional difference between conscious and unconscious mental processes. Instead, there is a material continuum out of which representational structures are composed, where the composition of thought-forms can be more or less transparent, and hence more or less susceptible to illumination, just as ordinary translucent objects can retain local variations in opacity. Conscious illumination takes place only with respect to a small portion of the causally efficacious material mind, and is not an intrinsic property of mental activities. Thus consciousness does not enjoy infallible or perfectly self-transparent access to the realm of content, nor does conscious presentation itself affect the functional role of the structures illuminated.

5 Conclusion

So, in a variety of respects, the ancient Sāṅkhya-Yoga version of dualism provides a more felicitous dividing line between substances than does the Cartesian parsing of mind and matter. Descartes' picture of consciousness and the mind has had a tremendous influence on the development of western thought; to a large extent it still defines the terms in which the mind/body problem is conceived, and it colors many contemporary notions of mental content and representation. In this sense, I think that the Sāṅkhya-Yoga view is at least as worthy of serious philosophical attention. A far wider range of cognitive phenomena are made available to naturalistic explanation on the Sāṅkhya-Yoga account, simply because the mind is included in the physical world. But, at the same time, the unique and autonomous status of conscious subjectivity is preserved. Of course, dualism of any flavor tends to be theoretically distasteful, and this is equally true of the Indian picture. Thus a primary motivation for examining this ancient philosophical system lies not in defining yet another type of polarity between substances, but rather in providing a different framework within which to consider issues such as the mind/body problem, which will in turn serve to articulate the dilemmas

that a satisfactory form of monism must address. By conceiving the problem along slightly different lines, the Sāṅkhya-Yoga view already solves some of the difficulties which would haunt a naturalized version of Descartes' model.

References

- Āraṇya, H. 1983. *Yoga Philosophy of Patañjali*. State University of New York Press; includes a translation of Patañjali's *Yoga Sūtra* and Vyāsa's commentary.
- Dasgupta, S. 1963. *A History of Indian Philosophy*. Vol. 1, Cambridge University Press.
- Block, N. 1978. 'Troubles with Functionalism.' *Perception and Cognition: Minnesota Studies in the Philosophy of Science*. Vol. IX, W. Savage (ed.), University of Minnesota Press.
- Feuerstein, G. 1980. *The Philosophy of Classical Yoga*. St. Martin's Press.
- Jackson, F. 1982. 'Epiphenomenal Qualia.' *Philosophical Quarterly* 32, 127–36.
- McDowell, J. 1986. 'Singular Thought and the Extent of Inner Space.' Appearing in P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*. Oxford University Press.
- Raju, P. T. 1985. *Structural Depths of Indian Thought*. State University of New York Press.
- Searle, J. 1980. 'Minds, Brains and Programs.' *Behavioral and Brain Sciences* 3, 417–24.
- _____. 1990. 'Consciousness, Explanatory Inversion, and Cognitive Science.' *Behavioral and Brain Sciences* 13, 585–96.
- Sinha, N. 1979. *The Sāṅkhya Philosophy*. Oriental Books Reprint Corporation, New Delhi; includes a translation of the *Sāṅkhya-Kārikā* of Iśvarakṛṣṇa.