Wolfram Hinzen

Dualism and the Atoms of Thought

Abstract: *Contemporary arguments for forms of psycho-physical dualism standardly depart from phenomenal aspects of consciousness ('what it is like' to have some particular conscious experience). Conceptual aspects of conscious experience, as opposed to phenomenal or visual/perceptual ones, are often taken to be within the scope of functionalist, reductionist, or physicalist theories. I argue that the particular conceptual structure of human consciousness makes this asymmetry unmotivated. The argument for a form of dualism defended here proceeds from the empirical premise that conceptual structure in a linguistic creature like us is a combinatorial and compositional system that implicates a distinction between simple and complex, or 'atomic' and 'molecular' concepts. The argument is that conceptual atoms, qua atoms, are irreducible to anything else. If so, and if the atoms are essentially semantic, a form of dualism follows: though positively inviting naturalistic inquiry into the semantic and mental aspects of nature, it requires that we look at the mental as a primitive domain of nature. Schematically, then, the argument is as follows:*

(1) Human consciousness/thought is conceptually structured.
(2) The human conceptual system is a 'particulate' system at a syntactic and semantic level of representation (the notion of a 'particulate' system is developed in Section 2).
(3) This implies the existence of conceptual 'particles', concepts that have no further semantic decomposition ('atoms').
(4) A conceptual atom cannot be explained in terms of anything that does not involve its own intrinsic properties (Section 3).
(5) Physicalism as normally conceived is inconsistent with (3) and (4) (Section 4).

Correspondence: W.Hinzen@uva.nl

1. Introduction

I shall introduce an argument, the 'Argument from Atomism' (AFA), which is an addition to a number of other arguments for psycho-physical dualism found in the literature,[1] though the reader should be warned that the notion of dualism used here may differ from one familiar usage of the term, where dualism is taken to be inconsistent with naturalism. The distinctive features of this argument are:

> (i) it does not mention 'qualia', or the phenomenal aspects of conscious-ness *per se*, focussing instead on aspects of consciousness that are con-ceptual and structural in nature (as well as, for this very reason, possibly unique to humans);[2]

> (ii) its essential premise is an *empirically plausible*, systems-theoretic hypothesis about what I call the 'particulate' character of the human linguistic system.

As a consequence of (ii), AFA is not an *a priori* argument for dualism that could claim to be conclusive, in the way that truths of logic or metaphysics might be claimed to be.[3] Indeed, my entire argument is based on the attitude of a broad *methodological naturalism*, and the paper as a whole strongly suggests the *need* for a naturalistic inquiry into the mental — though it makes the claim that the latter can, in a sense, only start off once the dualism proposed here is endorsed.[4] Note that it was precisely the present attitude of a methodological natural-ism — namely, the post-Aristotelian spirit of 'natural philosophy', and the ruthless physicalism and reductionism with which specifically Descartes aimed to implement it — which, in those early days, *entailed* a dualist consequence. More generally, naturalistic inquiry

---

[1] Meixner (2004) has a fairly comprehensive list of them, but it omits Bringsjord and Zenzen (2003).

[2] As for animal cognition, uncertainty extends, e.g., to whether non-humans will interpret the contents of their phenomenal consciousness conceptually (or have a concept of what they experience), and to what extent non-human animal thought reaches a propositional format. Propositions as expressed in language are structured entities, exploiting possibly language-specific structural resources. For two recent negative assessments of the ques-tion whether propositions are thought by non-humans, see Macphail (1998) and Terrace (2005).

[3] I make no claim here that metaphysical arguments for dualism are *im*possible.

[4] Otherwise there is in fact no domain for such an inquiry. There is still a myth that dualism is inconsistent with naturalistic inquiry in the sense of being inconsistent with the known laws or methodology of physics. This charge can be, and I think has been, adequately answered long since. Historically, it might be argued that ever since Newton, physics stopped providing any conceptual basis for materialism, a conclusion explicitly endorsed by Hume and Locke (see further Sect. 5). This is not to say, of course, that post-Newtonian physics provides a basis for dualism — though see Stapp (2005) and Barrett (2006) for dif-ferent arguments that quantum mechanics does suggest this metaphysical commitment.

simply leads where it leads. In Descartes, given the physics and the substance metaphysics of the day, it led to the insight that the world contains two substances. Today, I will argue, what we know about the mental aspects of nature still implies that they are radically different from and irreducible to non-mental ones, at least as regards their basic constituents (and possibly their structural principles too).[5]

Even the basis for dualism in the argument proposed here remains the one that is central in Descartes' argument for dualism in the *Discours* (1637), namely our human linguistic ability, as something that does not allow for a mechanistic or functionalist account (see Hinzen 2006a for more on this). Within the biolinguistic programme (Lenneberg, 1967), generative grammar (Chomsky, 1959; 2005) set out as an explanatory account of this linguistic ability, conceived in broadly Cartesian terms. The generative enterprise has later been largely understood as an instance of philosophical functionalism and the so-called 'Computational-Representational Theory of Mind' (see e.g. Fodor, 2000). As I have argued elsewhere (Hinzen, 2006a; Hinzen & Uriagereka, 2006), this is a mistake. The kinds of syntactic trees (see Section 2) that the generative tradition has been studying are something quite different from what many functionalist philosophers have taken them to be: as Hinzen and Uriagereka (2006) argue, they depict abstract hierarchical structures of concepts that, though correlating with brain structures, have no obvious physicalist definition. If functionalism is our basis for physicalism, it is, therefore (on this view) not only consciousness, but also language — as a cognitive system that provides specific hierarchical structures for our species-specific kind of consciousness — which causes cracks in the foundations of contemporary physicalistic opinion.

2. Language as a Particulate System

The human system of concepts that we express in language, I claim, is a *naturally occurring self-diversifying system* in the sense of Abler (1989). This section describes in detail what this claim means. Systems of this kind exhibit an unbounded variety of discrete elements,

[5] Perhaps we shouldn't call the 'mental' aspects of nature 'non-physical' today, since there is no reason to exclude them from 'physical nature' broadly understood as 'the only realm there is', or even from (a future) physical *theory*. After all, I assert dualism as a contingent truth within our current theoretical predicament and empirical knowledge of the mental, not as a truth of metaphysics (which some would want to distinguish from the truths of physical science). An ontological duality, that is, could exist even if mental entities *were* the subject of a future physical theory, a theory broad enough and appropriately revolutionized so as to account for the mental aspects of human beings. I return to these clarifications in Section 5.

and sustain this variation over time. All their productions are particulate units formed by a combinatorial operation that is *recursive*, in the sense that when the operation has constructed a unit, this very same operation can construct a yet larger unit by applying to the very unit it built in its first application. All complex synthetic (or 'molecular') units that are formed in this fashion bottom out, when fully analysed, with simple units ('atoms'), which have no (relevant) structural complexity. The identity of these atoms is unaffected by what compounds they enter into: they are freely *re*combinable. The complex units, equally, like chemical molecules, have a particulate identity, which again is preserved as such when they combine and recombine with yet other simple or complex units. The elements of any such productive system belong to a *generated set*, a set that can be determined by laying down a number of atoms and generative principles that combine the particulate entities of the system (both atomic and molecular) into more unboundedly complex ones. The successor function in arithmetic is a particularly simple generator of this kind: it generates an unbounded number of discrete units, the natural numbers.

Let me illustrate Abler's distinction in an abstract way using Figure 1, below. In (a), a combinatorial operation, +, merges two elements of the system, one black, one white, in a way that the information they contain is *averaged* into a shade of grey. Constituents *blend* and identities are *obliterated*. Consequently, variation will diminish over time. In (b), in contrast, identities are retrievable, and the constituents of the particulate composition in (b) can as such *re-occur* in other compounds (see (c)). Moreover, rather than arising from averaging out given properties (as in (a)), the properties of the formed whole are not *per se contained* in any of the constituents, being something 'emergent':
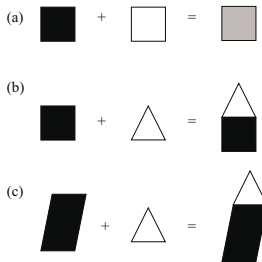

Figure 1

Temperature or the weather illustrate case (a): two temperatures combine into a blended average. Physical chemistry and the system of biological inheritance, by contrast, have a particulate character. In fact, it was Mendel's discovery of the particulate character of the units of inheritance — the genes — that saved Darwin's evolutionary theory from an inconsistency. If parents and offspring always combined into an average, evolution by natural selection would have less and less variety to make selections from, and would eventually come to a halt (as Fisher noted in 1930).

That human language is a particulate system too is the crucial premise in AFA. To begin with, human languages are widely recognized to be 'discretely infinite', in a sense that is most characteristically exemplified in the natural numbers, as noted above. As for infinity, just as there is no largest number, there is no longest sentence. For, in both cases, if there were one, one could use the very operations that built this object to construct a larger one. As for discreteness, just as each natural number is a discrete unit, each of the unbounded number of elements of a particular language is a discrete unit, which does not blend into any other as complexity builds up. Thus, we can construct an unbounded sequence of expressions, *loves Mary*, *John loves Mary*, *thinks John loves Mary*, *Bill thinks John loves Mary*, *Bill thinks John loves Mary without regretting it*, *Jill doubts Bill thinks John loves Mary without regretting it except on Fridays*, and so on, with each expression a discrete unit both syntactically and semantically. Similarly, it is not the case that for any two sentences, there would be a third sentence 'in between' them, in the sense in which there is an intermediate temperature for every two temperatures. If S is a sentence, 3/4 or 5/7 of S are not necessarily also sentences. Two sentences combined, like *John loves Mary* and *Mary loves John*, yield *John loves Mary and Mary loves John*, not some sort of average of both in which the original sentences do not figure any more as discrete units. This is an instance of the general fact that the facts aRb (a stands in relation R to b) and bRa combined yield aRb&bRa, in which the identities of all the constituents — namely, a, b, R, aRb and bRa — are preserved. 'Relational competence', in this sense, provides evidence for the *systematic* character of linguistic competence (Fodor and Pylyshyn, 1988), possibly a characteristic feature of the primate mind at large (McGonigle and Chalmers, in press; Gallistel, 1998).

Systematicity implies that the kind of relation that exists between aRb and bRa is not *statistical*, any more than the relation between x and y in the equation $y=x+2$ is statistical (cf. Marcus, 2001). Rather,

the idea is that whoever grasps an expression of the logico-grammatical form aRb, can grasp an expression of the form bRa — not because he or she happens to 'associate' one with the other, but for *structural* reasons, or as a matter of algebraic law. This is explained if that person's mind is structured such that he or she represents a, R, and b as simple constituents of the complex mental representation in question, plus the structural fact that the relation in question holds between *a* in its *first* argument position, and *b* in its *second*. For then, *ipso facto*, an expression can be represented in which *R* holds between *a* in its *second* argument position, and *b* in its first; this is so if *a* and *b* have their respective identities *independently* of which argument position they figure in, and if the argument positions are the ones they are *independently* of which arguments figure in them. We may thus view these argument positions as *variables* which take *values*, where crucially the variables do not change depending on what values they take, just as the variable 'x' does not change its algebraic role in the equation y=x+2 depending on what number provides its value).[6]

Importantly, the systematicity of language is not limited to the *syntactic* properties of expressions. Languages are systematic in the sense, not only that the expression *John loves Mary* can be syntactically represented just in case *Mary loves John* can, but also in the sense that the one expression is semantically *understood* when the other is. This *semantic systematicity* is again explained if the same constituents figure in the same argument positions, though in reverse order, where being the 'same' constituent now entails having the same meaning. This very assumption is also needed to explain the systematicity of *inference*, since it is because the expression *turtles* has exactly the same meaning in the sentences *Rabbits are faster than turtles*, *Ferraris are faster than rabbits*, and *Ferraris are faster than turtles*, that the last of these expression is a valid deductive consequence of the former two.

Like other particulate systems, languages *sustain* their variation. Contrary to a still widespread prejudice, natural languages do not for example linearly increase or decrease in their structural complexity as a function of the cultural sophistication of their speakers. There are no 'primitive' natural languages. Also, in language contact situations,

---

[6] The systematicity of thought and language is a challenge to the cognitive science community that Fodor and Pylyshyn formulated in 1988. Their claim was that as a consequence of this feature, *syntactically structured mental representations* must figure *in a causal role* in the mind, the current lack of neurological evidence for such representations notwithstanding. Hadley (2004) surveys recent attempts to meet this challenge, and argues that, despite many claims to the contrary, it is far from having been met, especially if systematicity in the sense of the following paragraph is taken into account.

languages mix to form a new language, but they do not do so in such a way that they become an average of the contact languages involved, or that for every two languages there develops a third one in between (in which case variation would be continuous, which it isn't, as Baker, 2001, p. 82 points out). Rather, new languages arise in the way that all human languages are built from the same particulate resources, by *recombining* their distinctive elements (Mufwene, 2001).

Consider now further that in the process of building complex expressions, words don't *blend*. If we put the lexical items *kill* and *Bill* together, the verb phrase [*kill Bill*]$_{VP}$ forms, where square brackets indicate (acoustically invisible, but mentally represented) phrase structure boundaries enclosing a higher-order unit of linguistic structure. This unit is not interpreted as some kind of average of *killing* and *Bill*, but as an *integrated event* that has an action in it that has intrinsically a participant in that action, in linguistic terms the 'Patient', which is the person who suffers the action in question. It is, moreover, universally the *first* argument of a transitive verb that figures as the Patient (or 'Theme') in the relevant event, never as the 'Agent'; and note also that the event has an internal *implicational structure*, in that it implicates (or is inherently *bounded* by) the state of Bill's being dead, a state that it has as its intrinsic end or 'telos'.

These properties of the VP can be no more predicted from those of *kill* and *Bill* than the properties of H_2O (e.g., its fire extinguishing properties) can be predicted from those of H_2 and O (the latter of which is fire-enhancing), or the properties of a gene can be predicted from the nucleotides that make it up. Again, then, these are *emergent* (qualitatively different) properties, not predictable from or contained *per se* in those of their parts. Forming phrases and sentences also importantly involves a relation of *predication* between two constituents, which is nothing as trivial as *associating* the verb and the noun, or *concatenating* them, or first *thinking* 'kill' and then thinking 'Bill' (Macphail, 1998, Ch. 5; Fodor, 2003).[7]

Crucially, the complex event representation [*kill Bill*]$_{VP}$ does not relate to its constituents in the way that the product *6* relates to a possible factorization of it, say the factors *2* and *3*.[8] The reason is that this kind of relation would leave the systematicity of language, which needs explaining, unexplained: If *kill* and *Bill* were related to the complex event representation [*kill Bill*]$_{VP}$ as *2* and *3* are to *6*, the

[7]   What precisely predication is — hence what provides for a solution to the traditional problem of the unity of the proposition or the phrase — I take to be a crucial and unresolved problem.

[8]   As Fodor and McLaughlin (1995, pp. 212–14) pointed out against Smolensky (1995).

expressions *Bill killed Bill* and *Hill killed Bill* would not be systematically related to one another. For, the above explanation of this relation depended on the fact that *Bill* is a (context-independent, syntactic and semantic) *constituent* of these expressions, in the sense that it is *necessarily tokened when these complex expressions themselves are*, and thus can be a *cause* (or can explain) their lawful relationship. But an element in a possible factorization of *6* is *not* necessarily tokened when *6* is, and does not govern its causal properties. That *6* can be factorized does not make it structurally or syntactically complex in the way that [*kill Bill*]$_{VP}$ is.

Just as words don't blend when combining, phrases don't either. Thus, we may separately assemble two phrases [John's wife] and [likes this], and combine them to obtain the new mental representation [[John's wife] [likes this]] (Figure 2):
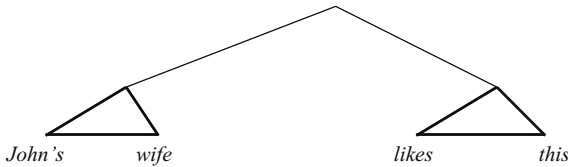

Figure 2. Phrases combine with others as whole units to form new phrases (a sentence in this case).

In this representation, [likes this]$_{VP}$, though structurally complex, functions as a particulate unit when interacting with the other phrase [John's wife]$_{NP}$. We can then move *this* to the front for emphasis, to obtain *This, John's wife likes*. On standard current accounts, the result of this operation is a 'discontinuous' relation between two elements in different positions of a phrase marker, namely the moved *this* and its original occurrence, which together form a so-called *chain* (Figure 3). But the fronted *this* will now crucially still be interpreted as the object of *likes*, and as the Theme of the liking. Hence the information that *this* is the *internal* argument of *likes* (or part of the Verb Phrase, VP) is *preserved*, a fact that classical generative theory symbolizes by positing a 'trace', t, in place of the launching site of this syntactic transformation:
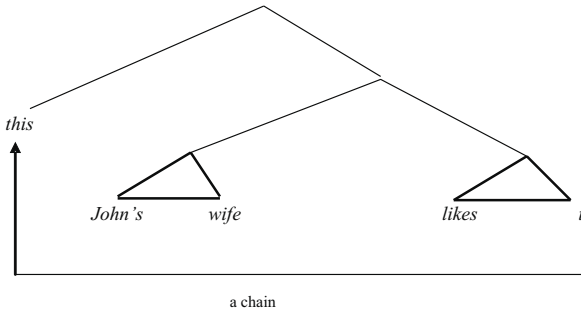
a chain

Figure 3. Transformations of a given tree structure (chain formation) pre-
serves all original hierarchical information by leaving 'traces'.

If, as in recent grammatical theory, the trace is interpreted as simply an
identical *copy* of the moved item, what we thus notice in the object
above is that the movement of *this* has left the previous object com-
pletely unchanged: it is a particulate unit whose internal structure is
opaque to further transformations. Indeed, if we try to 'tamper' with
what the original phrasal constituents were, and try inserting 'this'
into the other phrase [John's wife]$_{NP}$, ungrammaticality results:
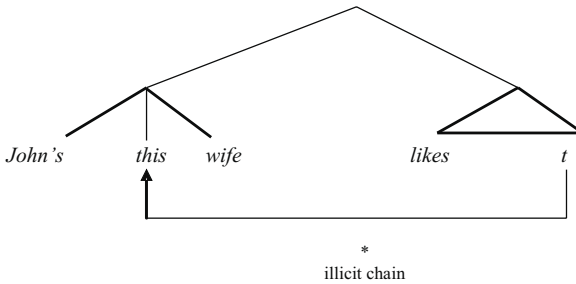

*
illicit chain

Figure 4. Transformations of given phrase structure cannot override
given phrase structural units once these are formed.

In short, human languages do not just have particles at the lowest level
(the level of the lexicon, or of words), but at *several* levels. By con-
trast, though grains of sand combine into hills or dunes, these are not
particles which would combine again systematically with other such

particles at their own level of organization to form yet other and larger particles.[9]

Now, particulate structure at several levels of organization is systematically important for the way that syntactic structures are mapped into semantic structures. It is because *John loves Mary*, at a syntactic level of representation, *contains* the constituent or particle [loves Mary]$_{VP}$, that it is not interpreted as *Mary loves John* is, which contains *no* such syntactic constituent (instead, it contains the constituent loves John][$_{VP}$). This exemplifies that syntactic form is mapped *transparently* into semantic form — syntactic constituents *correspond* to semantic ones. Transparency in this mapping would also predict that *kill* does not and cannot mean what *cause to die* means, simply because *kill*, contrary to *cause to die*, does not have *cause* or *die* as constituents, even if killing necessarily implicates the causing of death. This prediction is born out in the following ways, among others. Whereas Bill *could at noon cause Hill to die* at midnight (say by tampering with his telephone and having someone give him a ring at midnight, causing an explosion), we couldn't felicitously comment on this by saying Bill *killed* Hill at noon. And whereas, if we say that *Bill caused Hill to die by pulling the trigger of his pistol*, Hill could possibly have died by pulling the trigger of his own (Hill's) pistol, it is the case that, if we say that *Bill killed Hill by pulling the trigger of his pistol*, Bill will have pulled the trigger of his own (Bill's) pistol (the other interpretation becomes ungrammatical).

This is evidence that a word like *kill* is not only syntactically but also semantically *simple* — it is an atom — contrary to a phrase like *cause to die*, which does exhibit syntactic and semantic structure. Put differently, morphemic boundaries *matter* for semantic interpretation. Something that does not *seem* to have constituents, like *kill*, does not have them covertly, or at some 'semantic level of representation', either. There is a *necessary* relation between killing and causing death for sure, but it is not a *structural* one, or vindicated by the constituent structure of *kill*. For all of these reasons, *kill* functions as an atom semantically, in contrast to phrases.

I will assume in what follows that this architectural principle, the *Transparency of the syntax-semantics mapping*, is substantially correct. Indeed it has been empirically very fruitful, and is in any case the

---

[9] Phrases, when combining, either generate more phrases, or else generate chains, depending on whether a phrase, B, with which a phrase, A, combines, is *external* to A prior to merging with it (as in Figure 2), or is a *constituent* of A, and hence *internal* to it (as in Figure 3). Technically, the former process is referred to as 'external Merge', the latter as 'internal Merge'.

strongest hypothesis: one would go for a non-transparent syntax–semantics interface only if forced.[10] Note also that even fervent anti-atomists like Jackendoff (2002) would accept the thesis that the human conceptual system is a combinatorial and particulate one that bottoms out with non-complex conceptual atoms.[11] At this point, therefore, I ask readers unconvinced by lexical atomism or the particulate character of human language simply to grant this character and hence the existence of conceptual atoms as a premise in the argument to come.

3. The Essence of Atoms

Let us now ask: what makes a particular conceptual atom the atom it is? Which of its properties does it have *essentially*, in the sense that it could not lose them, while remaining the atom it is?

Clearly, the *content* of any concept C is essential to it, *whatever* notion of content one has (e.g., a referential or a use-theoretical one): concepts are *semantically individuated*. On *anyone's* notion of meaning, I submit, it would seem that without meaning *house*, the concept of a house would not be the concept it is. I here take it as undeniable that, again whatever one's notion of meaning, one will concede that the concept HOUSE means *house* and not HORSE, say, CHAIR, or PICCADILLY.

However, if we say that what makes the concept HOUSE the concept it is, is that it means *house*, we are not talking informatively, or non-circularly, since the concept itself is made use of in specifying what is essential to it. Thus the question arises: can we specify any concept C in *other* than *circular* terms? In particular, are there properties essential to it, other than its content as circularly specified? This question will occupy us for the rest of this section, and the negative answer I will give for the case of conceptual *atoms* will proceed by an elimination of available philosophical options in the theory of conceptual content. Before starting I note that this negative answer is actually quite expected: if the explanation of the meaning of a complex concept involves an appeal to the meaning of its (ultimately) atomic parts (and it does, if compositionality and particularity hold), then for these parts *another* kind of explanation must be found. But it is thoroughly unclear — and certainly unknown — what that explanation should be.

---

[10] For recent defences of lexical atomism and this transparency claim, see Larson and Segal (1995), Borer (2004/5), Uriagereka (forthcoming), Fodor and Lepore (2002).

[11] Jackendoff would not equate these with the meanings of specific words, but in practice they are still identified by means of ordinary words, such as PATH, MOTION, and CAUSE.

3.1 Appeal to representations

Let us first discuss *representations of* some arbitrary conceptual atom
C as a possible candidate for what is *essential* to C. To illustrate how
this (generally speaking ill-defined)[12] notion of a representation has
been used in philosophical discussions, suppose you are thinking that
the sky is blue. Then that's the content of your thought, or what the
thought is 'about'. It might then be suggested that for you to think this
thought with this content, there must be a *representation* of that con-
tent in the brain. The thought itself however is not what's in the brain,
it's merely what's being represented by something in the brain. The
advantage of the stipulation is that we can explain thinking as a pro-
cess: it will simply be a computation of these representations, which
may have some sort of symbolic or quasi-linguistic character. So, on
this view the relation between a thought and a representation is like
that between what is a meaning and what has a meaning; in the same
way, a symbol we introduce to denote a thing has a meaning, but isn't
one.

Now, any such particular representation R_C of the atom C may be
syntactically or *semantically* individuated. In the former case, we
describe it in purely physical or physiological terms, i.e. without ref-
erence to any meaning or concept (for a *functional* individuations see
subsection 3.4). In the latter case, the question to be asked immedi-
ately is: in virtue of *which* semantic content is C individuated? Again,
that shouldn't be the content of C itself, on pain of circularity. But can
concept(s) *other* than C individuate R_C in a way that we know it is the
representation *of C*, and of no other concept? We will address this
question in subsection 3.2. Suppose then for now the first option, that
the representation R_C of some conceptual atom C is 'syntactically'
individuated, i.e. without reference to meaning properties. Then why
should it be essential to C how it is so represented? Clearly, if R_C is a
physical or physiologically described object, then R_C as such, on
standard physicalist views, has *no* semantic properties at all. If C has
its semantic properties essentially, as we stated, R_C can't be essential
to C.

If we shift from neural representations to linguistic ones, this point
becomes obvious: it is irrelevant to the concept TREE whether the
acoustic pattern \arbre\, \tree\, or \baum\ represents it. Similarly, it

---

[12] The following illustration is meant to explain the use of the term representation in the his-
tory of analytic philosophy. As will be seen, it in particular depends on a distinction
between 'syntax' (form) and 'semantics' (content) which I consider obsolete (see e.g.,
Hinzen 2006a). If the notion is dropped as empirically irrelevant, the attempt for a further
analysis of meaning that this sub-section illustrates becomes redundant.

would be obscure what it would be about some psycho*physical* property of the brain, call it P10, that makes it have the content *tree* more than some other psychophysical pattern P11, or that makes it mean *tree* more than *house*, or *person*. Though I assume that psychophysical *correlates* of C will exist, they don't explain why the concept correlated with it means what it does. Whatever relation there is between a meaning/concept and its (psycho-) physical representation or correlate, it seems it will not hold because of any essential or semantic connection between the two; quite the contrary, the connection is thoroughly *contingent*.[13]

We thereby reach our *first* conclusion: As long as its *content* is essential to a conceptual atom, its (physically described) *representation* cannot be. Basically, to switch from talking about conceptual atoms to talk about their representations is to change the subject.

3.2 Appeal to relations of concepts to other concepts

A more promising place to look for things that may be essential to an atom C is relations of C (as *semantically* individuated) to other concepts. But there appears to be no known case where such relations literally *constitute* an atomic concept. Consider e.g. the concept of BELIEF. There is no known way to relate this concept to other concepts in a way that possessing these other concepts and some means of combination of them would account for possessing or acquiring the concept BELIEF. As Leslie (2000) explains in painstaking detail by analysing a number of available proposals, either the other concepts in whatever combination are *weaker* than the concept of belief (in the sense of not entailing it), or they are *at least as* strong. Either way, they do not account for it. The concept of a belief is simply not known to follow from anything other than itself.[14]

---

[13] Research on the neural correlates of consciousness does not seem to state more than brute and ultimately unexplained psycho-physical correlations either (see e.g. Dehaene, 2002, for a recent discussion of arithmetic, or consider Koch's [2004, p. 2] concession that it remains a mystery how meaning should arise from the electrical activity of the brain). For 'semantics by stipulation' in regard to connectionist networks see Fodor & Lepore (2002, pp. 161–5).

[14] I think that the same conclusion can be reached by looking carefully at available proposals on how concepts like (natural) NUMBER are acquired by the child (see e.g. Carey, 2004). One would expect that the child does so on the basis of strictly *weaker* concepts; but then the *transition* to NUMBER would be unexplained. Or the concepts that it starts with are *equally* strong, in which case the account is circular again. Carey claims to solve this problem via the notion of *bootstrapping*, but bootstrapping is a process of mapping between *given* (semantically individuated) representations. It does not make the representational resources of the mind *strictly stronger* than they (implicitly, prior to the explicit mapping)

It seems that conceptual atoms are *ipso facto* not individuated by relations to other atoms that they co-occur with: that's why they are atoms. To be an atom is to be (syntactically and semantically) structureless. If relations to other concepts *were* essential to one such atom, they *would* be part of its structure; the other concepts would *figure* in that very atom, just as, if the concept KILL *was* essentially related to the concepts CAUSE and DIE, the concept KILL would literally *be* the concept CAUSE TO DIE, and thus the concept CAUSE would *figure* in it as a constituent. Recall also that our explanation of systematicity demanded that the constituents of syntactically structured expressions have their meanings in a *context-independent* fashion, i.e. independently of what other concepts they co-occur with, or what molecular compound they enter into (see again Hadley, 2004, for discussion of this point). *House* does not mean what it does because we can plug this word into an argument slot of, say, the expression *John loves houses*. On the contrary, *John loves houses* means what it does because it contains the constituent *houses*, which in turn contains *house*, which context-independently means HOUSE.

None of this, I hasten to add, is to deny the intuition that there are numerous, and even *necessary* relations between conceptual atoms and other concepts. But it simply doesn't follow, from it being a necessary entailment that Bill is dead if he has been killed, or that Bill has been caused to die whenever he has been killed, that there is any *semantic* or *conceptual* connection between the concept of killing and that of (causing) death. Evidence that there is such a conceptual connection would be that one couldn't conceivably think the one concept without thinking the other (just as, indeed, one cannot think the concept *kill Bill* without thinking *Bill*). But that one cannot do this is extremely implausible. Consider this dialogue:

> You: 'Hi there, didn't you just kill somebody?'
>
> A: 'Yes.'
>
> You: 'Isn't there something you *did* to accomplish this?'
>
> A: 'Yes.'
>
> You: 'How would you describe this?'
>
> A: 'I killed him.'
>
> You: 'Yes, but *how* did this happen?'
>
> A: 'Well, I decided to kill him, I strangulated him, and then he was gone.'

---

were. It is not clear at this point how our concept of number, with the infinitary property it has, could follow from anything other than itself.

You: 'What do you mean? Wasn't he, in the end, …'

A: '… *gone*? … *killed*?'

You: 'Hm. Never mind. And when you yesterday *boiled the beans*, don't you think there was something in *common* between what you did today and what you did yesterday? You did *some*thing, … X …, and then something got to be the case as a result of that.'

A: 'Huh…? Well, yesterday I boiled the beans, and they became boil*ed*, and today I killed someone, and then he became kill*ed*.'

You: 'Hm, hm. But look, suppose your victim today had fallen over a stone instead and got … *killed*, as you put it, without *your* having done anything to *accomplish* this. What would be *different* in this case?'

A: 'I wouldn't have killed him.'

You: 'Right, right … But the stone clearly didn't *kill* him, it rather merely *did* something, X, so as to make him *be* in a lifeless state, Y. How would you call what he did?'

A: 'Made him lifeless, perhaps?'

You: 'Rats. And when you killed this poor fellow, what would you say is the relation between *you*, and *him*, that had the *consequence* of his being killed?'

A: 'I am his *killer*, of course!'

I think this conservation could continue for a very long time, without A ever hitting upon or using either the concepts CAUSE or DEATH, despite your best efforts to trigger these concepts, if, for some (perhaps neurophysiological) reason, he simply *lacked* these concepts. This is evidence that one can lack the concepts CAUSE and DEATH, and yet have KILL, and shows that the former are neither constituents of the latter, nor essential to it in our required sense. Similarly, it is plausibly the case that despite the *mathematical necessity* that the number 2 is the only even prime, there is no *conceptual* necessity to that. It is simply untrue, it seems, that, by having a thought about 2, we thereby have thoughts about primes (a belief ascription involving the concept TWO could be felicitous in circumstances where a belief ascription involving PRIME clearly would not be). See Fodor (2004) for discussion.

The *second* conclusion we are thus envisaging is that relations of C to other concepts, while obviously existing, are also non-essential to it. Traditionally, this same conclusion has been drawn by appeal to the breakdown of the analytic-synthetic distinction at the level of the

lexicon. While there arguably *are* analytic inferences where we have a *structured* expression — thus, arguably, it follows from the grammatical structure of the linguistic expression[15] *brown cow*, and not our beliefs about the world, that a *brown cow* is *brown* — there are no analytic inferences where we have a structureless lexical atom. Except for the property of being a cow, there simply is no known property that we think a cow has, the lack of a possession of which would lead us to judge that we are *not* dealing with a cow, and the possession of which would necessarily lead us to call it a cow. Clearly, e.g., we need not agree that a creature whose *genome* was not like that of standard cows, would therefore not be a cow (in fact we could comment: 'That cow is a rather nonstandard one!', in which case we think of it as a cow, though a non-standard one).

For a final example that will concern us again later on, it seems rather plausible that when you think *London is polluted*, *I enjoy London*, *London is very green*, *London is corrupt*, *No city is noisier than London*, and *London is a city*, you are having six thoughts *about London*, involving reference to the very same city in the UK. But clearly, none of these beliefs, and none of the inferential connections involved between *London* and all the other words above are analytic; if they were, the lexical item *London* would, when entering a syntactic compound, take these beliefs and inferential connections with it when contributing its meaning to the determination of that of the compound. But it doesn't, since if it did, *London is not noisy at all* would be self-contradictory, as would *London is not a city*. But neither is self-contradictory; the former not if one is deaf, say, or happens to find London peaceful; and the latter not if, say, London becomes an *island* rather than a city in our judgement because Europe is flooded by a Tsunami and only a few houses are left of it, making up a village on an island.[16]

---

[15] The notion of a linguistic expression here is the linguistic, not the philosophical one: an expression is a pairing of a particular sound with a particular meaning. See e.g. Chomsky (2000).

[16] One of my anonymous referees has the inclination to say that in that case 'London had ceased to exist, rather than that it still existed but was no longer a city'. But all I am saying is that this conclusion is not necessary, i.e. that it is conceivable that a single sane person will find the assertion that London, though not a city any more under these circumstances, still exists. If the referee is right, any such assertion would be a conceptual contradiction — contrary to fact, I should think. *Ipso facto*, CITY is not an essential constituent of LONDON. It *clearly* seems that in talking about cities we often take their histories into account, hence what makes London London, need not be that it is a city, but that it has a certain continuity over time.

3.3 Appeal to 'reference'

Let us then return a second time to the question how and whether we can specify what is essential to an atom in non-circular terms, and try it a third way. This way is to declare that mind-world, or *referential* relations are essential to an atom C. Again, a commitment here must be that this reference-relation can be identified *without appeal* to the concept in question, or a closely related one. Is this possible?

Not, say, in the London examples we gave above. Here we were referring, in one sense, to radically *different* things in each of the six initial sentences: to the air quality in a certain place in the first sentence; to the place's entertainment value in the second; to its colour in the third; to its business practices in the fourth; to acoustic properties in the fifth; and so on. At the same time, in each case we were referring to nothing other than *London*, i.e. the very *same* city. So in one sense the reference-relation is entirely stable; in another it shifts as our perspectives and predications do. Now, can we specify *what* remains stable — hence what alone can qualify as being *essential* to whatever meaning our word *London* has in the first place — without appealing to that meaning? It seems not. As we noted, it is only a fact, not a conceptual necessity, that London is a city; in fact, it is not even essentially a geographical *place* (it can be destroyed and rebuilt 100 miles away, and we might still think of it as *London*, commenting that 'London is now not any more on the Thames', as Chomsky, 2000, emphasizes). Of course, what remains stable also cannot be specified in *physical* or non-intentional terms, for, *physically*, London does *not* remain London, of course, if its location, colour, or noise change. It is because we have the concept of a city that we do, that London may well remain itself under certain physical changes, which shows the concepts are not determined by physical parameters of the object referred to.[17]

Let us also formulate a slightly different objection. One natural way to think of reference is to imagine a person using the word *snake* to refer to snakes. But such a person *possesses* the concept of a snake; she is thinking of the object of reference *as a snake*, hence her act of reference involves a perspective, predication, and description, as in the cases of *London* above. Reference to snakes spelled out in merely causal terms would not be the same thing at all. Thus we might build a snake detector that would reliably sound an alarm if a snake came along, all in the absence of a possession of a concept of a snake, or

---

[17] This verdict against the explanatory use of positing a 'reference-relation' is essentially Chomsky's, and developed from there (see Chomsky, 2000, and Hinzen, 2006c).

indeed any concept at all. 'Reference' as performed by non-human animals might well universally be like that: a conditioned causal relationship, in the absence of intentional meaning and reference.[18]

If a causally specified relation of reference would be consistent with the *absence* of any snake concept, it does not look as if causal relations to the world alone could possibly specify any concept. The same logical independence holds the other way round: we can also retain a concept, while physical or causal relations it is associated with disappear. Thus we may have a concept of a person, say Jack, whom we know, and that person could change its physical properties — which alone could figure in or determine a causally specified relation of reference — while *remaining* the person he is (say, he could lose a leg, or replace his heart). Indeed he could, in our judgement, lose *all* his physical properties, and without any difficulty we could find and comment that *Jack is now in heaven and happy*, an expression in which we are referring to exactly the same person, though there is no physical referent any more.[19] We might equally find out, or come to believe, again without any *conceptual* difficulty involved, that Jack was never more than a *soul* that was sent telepathically into our minds, and that was withdrawn afterwards by those who sent him; in which case he not only lost his physical properties, but never *had* them, without any of this necessarily affecting our acts of reference to *him*.

I cannot see how, in the light of this, *non-circularly (e.g. causally) specified* referential relations could possibly be rated *essential* to a concept; it seems, on the contrary, that the kinds of external acts of reference that we engage in are *conditioned by* our possession of the relevant concepts, and hence do not *explain* them. It is because our word *Jack* expresses a person concept, and because it is a feature of our specific human concept of a person that embodiment is not conceptually necessary for being a person, that he can remain the person he is if he loses his body (same for the London case).[20] How would one decide the question of whether one was referring to the same thing, if not by

---

[18] Terrace (2005) summarizes five decades of ape language research by saying that there is '[n]o evidence that apes used any of the symbols they learned to *refer* to objects or events, or that those symbols had any function other than to request food or drink'.

[19] One anonymous referee finds such examples 'exotic', but we are not talking about what is worthy of belief here, rather about what is *conceptually* possible or impossible. There is no other way to find this out than by stretching our concepts as far as we can, in an effort to see which situations they would still adequately describe. That bodiless individuals are in fact *routinely* conceptualized in the absence of any feeling of conceptual contradiction whatsoever should be clear from turning on TV late at night, and looking for the right movie genre.

[20] Though, again, I am not saying that every speaker would *necessarily* make such a judgement.

consulting the concept we have of Jack? Nothing else would tell (see further Hinzen, 2006c).[21]

This then is my *third* negative conclusion, that causal-referential relations, though clearly involved in human acts of reference in which an atom C figures, cannot be *essentially* involved in such acts either. If the concept is a *precondition* for such acts to be what they are in the human case, using them in these acts can't be a precondition for having them.

3.4 Appeal to relations between representations

Let us finally look at the option that what it is to be C *is* spelled out relationally, but now by appeal to relations *between* (non-semantically individuated) representations (including possibly an appeal to a speaker's 'dispositions' to relate such representations to one another), in the sense of section 3.1. But if our ontology of concepts consists in representations and (causal) relations between them alone, then it is in fact not clear why this should be offered as a solution to the problem we started with in this section: the question of what makes a concept the concept it is, and how what is essential to it can be spelled out in non-circular terms. The option we now envisage has at least *prima facie* nothing to say about concepts; it does without them, since only (non-semantically individuated) *representations* are involved. On our earlier linguistic analogy, it would be like talking about sound-patterns alone, in abstraction from semantic patterns that these correlate with. From such sound patterns no semanticity ensues. The child learning a language doesn't bootstrap its semantic knowledge from phonetic, acoustic, and perceptual data: in fact, there is good empirical evidence that it *uses* its prior semantic and syntactic knowledge in order to find out *which* correlation obtains in its environment between the sounds it hears and concepts that figure in its thinking (see Gleitman *et al.*, 2005).

It is thus again unclear how semanticity should be bootstrapped from the non-semantic, such as, on the option under consideration, manipulations of physical symbols or dispositions to engage in such manipulations. If however we loosen the physicalist constraint that

---

[21] The same anti-referentialist conclusion follows directly if we look at concepts that children have of abstract algebraic relationships between words, e.g. phrase structural relations. The child's concept of a phrase structure boundary cannot be explained by a relation of reference between this concept and phrase structure boundaries as such, because phrase structure boundaries, qua objective, mind-independent entities, don't exist. Non-human animals equipped with very similar sensory machinery do not detect them when exposed to them. They are in the mind of the (human) beholder, or require the language faculty, and simply are not there at a purely acoustic or physical level of description.

the relation of concepts to physical symbols and our behavioural manipulations of them must be explanatory for their content, and specify these behavioural manipulations in semantic or intentional terms, the circularity problem is back on the scene.

Suppose we wish to engage in some 'inferential role' semantics for the concept CONJUNCTION, without of course invoking the content of the word *conjunction*. Note that if an inferential semantics works for anything, it should work in this case, since for this concept there *are* at least known inferential rules that with some plausibility exhaustively capture its logical or Boolean content: the Gentzen introduction and elimination rules (e.g., 'from A&B to infer A and to infer B'). But, as Fodor (2004, p. 45) points out, understanding these rules of inference (recognizing their validity) *depends* on a grasp of the meaning of '&'. It does not merely depend on an appreciation of the intricate physical shape of this symbol, say. It's that very grasp that must *explain* why this inference and no other inference is drawn (it shouldn't be drawn merely by means of some sort of physical reflex, or a drug, if the relevant agent is to be credited with a possession of the concept CONJUNCTION). In short, the Gentzen-rules may well specify what CONJUNCTION means, but *for someone* who already *has* that concept. They explain what it *is* to have it.

A second issue is that the account now under considerations seems inconsistent with the *particulate* (and ultimately atomic) character of the human conceptual system. For a concept to consist in causal and inferential relations between representations, it has to be *complex*: the statement of its possession condition would mention other concepts, hence this proposal can't apply to atoms.

A third problem is that we know on formal-logical grounds, as Hadley (2004) points out against connectionist reductions of concepts, that *no* manipulations of representations can possibly implicitly define any specific semantic content: no axiomatization of such symbol manipulation, e.g. through Peano's axioms of arithmetic, will determine the content of our ordinary number concept uniquely.[22]

---

[22] This is due to non-standard (non-isomorphic) models of axiomatized languages that will necessarily exist (at least for first-order languages), and in which such an axiomatization would still be true. Penrose's related Gödelian arguments to the effect that such formal languages cannot ever *exhaust* the content of, say, our number concept (Penrose, 1994), have of course been heavily criticized, especially by Bringsjord & Zenzen (2003, Ch. 2). But even Bringsjord & Zenzen's arguments do not establish anything against Penrose's non-computationalist *conclusion* (for which the authors indeed provide a more convincing argument themselves), and it is widely agreed at least that the formalist reduction of mathematics that Hilbert attempted has been deadened after Gödel.

Fourth and finally, inferential semantics is in a well-known conflict with the compositionality of language that is not easily resolved: if meaning is compositional, and inferential roles are not, then inferential roles can't be meanings.[23]

3.5 Only the content is essential

Having reached this, our *fourth*, negative conclusion — that inferential relations among representations are not in general meaning-constitutive either — our search for ways in which what *is* essential to an atom, namely its content, could be specified, described, or explained, in non-circular terms, has yielded no positive result. The *fifth* conclusion we are thus facing in this section is that — on the empirical assumption of the absence of further options — *only* its content is essential to a concept.[24] The only way to spell out what is essential to a concept like HOUSE is to say that it means *house*. There is no *non-circular way of spelling out a particular concept*, if it is an atom. For this to be so, it would have to essentially relate to something else than itself. This is *not* so for atomic concepts, though it *is* so for complex concepts: thus, the meaning of BROWN COW *can* be derived without appealing to that very concept and its content, for on the assumption of compositionality it follows directly from the meaning of BROWN, COW, and its syntax.

That a concept that is an atom has no relational specification also has the consequence that it is not *functionalizable*. Its being functionalizable would mean we could define it by appeal to a particular causal role that it intrinsically plays; but if the above is right, the *right* causal role, if there is one with which the concept can be identified, is not identifiable other than circularly, that is by reference to the concept. In an explanation of the form: 'A physical symbol S plays

[23] Compositionality entails that the meaning of BROWN COW is determined by that of BROWN, of COW, and the syntax of this representation. But the inferential role of BROWN COW in your thought might license an inference to DANGER, and that aspect of the meaning of BROWN COW is not derived from that of its constituents plus syntax, hence is not compositionally determined and is non-intrinsic to this complex representation (see Fodor & Lepore, 2002, ch. 1).

[24] A referee remarks, correctly in principle, that from the fact that several determined and ingenious attempts to devise viable physicalistic explanations of semantic atoms have failed, it does not follow that it is impossible that any such explanation will ever be discovered. But the intrinsic semanticity and undivisibility (structurelessness) of atoms makes this hard to see (though, as I emphasized, the present argument is not a deductive, *a priori*, or conclusive one). If it is then argued that this epistemic point does not matter, Section 5 should be considered, where it is argued that even if such an explanation is discovered, that explanation should be consistent with the fact that semantic properties exist as much as non-semantic properties do, with no reason to reduce one to the other.

such and such a causal role iff it means …' would have to evoke a reference to the intended meaning on the right hand side. Functionalizable, indeed, is only what is *computable*; but computable is only what has structure; and atoms have no structure, by definition. Functionalism as an account of semantic content fails for *atoms*, even if nowhere else.

It is interesting that Fodor — known also as a functionalist — reaches what essentially amounts to the above conclusion in Fodor (2004), as Rey (2004) notes with some uneasiness, when asking how Fodor's recent views still differ from those of some philosophical 'mysterians' and dualists like Bealer (2002). There aren't, Fodor argues in Fodor (2004), '*any epistemic clauses in the analysis of concept possession*' (p. 32), where any such epistemic clause would constrain possession of a concept C by the having of certain epistemic capacities, say to be able to sort Cs from non-Cs, or to draw inferences in which C figures. But, as Fodor plausibly argues, the question which *kind* of sorting would substitute for possession of a concept C can ultimately only be answered by letting C *enter into* the description of the sorting, which means the sorting can't be essential to C or explain it. E.g., if the concept TWO is the concept whose possession we wish to specify, then not even a sorting according to a necessarily equivalent concept like THE ONLY EVEN PRIME will do, for a person can sort according to the one, but not the other. Yet, the person's sorting will overtly or physically look the same in both cases. What sorting then *will* do to specify possession of the concept TWO uniquely? Only a sorting according to the concept TWO, Fodor argues (p. 39), which is precisely the circularity I argued for above.

4. Consequence of the Essence of Atoms

The above fifth conclusion and its corollaries imply a form of dualism, I now claim. For, atoms,

    A   if being essentially semantic,
    B   and not relationally specifiable,
    C   are not entailed by the material
         (if by the 'material', as customary in current discussions,
         we mean the non-semantic).

Accordingly, our atoms are not only primitives, they are *non-material primitives*, reducible to nothing else that would be of a material nature.

This argument doesn't imply anything against a programme such as that of searching for the neural correlates of consciousness (NCC)

(see e.g., Changeux, 2004; Koch 2004). But it *constrains* that programme: firstly because whatever NCC of conceptual thought there will be, it will have to make sense of the existence of conceptual 'particles'; and secondly because there is no way, if the above argument is correct, to identify an atomic concept with a neural cell assembly or let it be supervenient upon it. For the latter, but not the former, is complex, non-semantic, and functionalizable, while concepts are simple, essentially semantic, and non-functionalizable.[25] I will now describe in more detail what I take the derived dualism to actually mean.

5. The Meaning of Dualism

Quite possibly, materialism is less a false doctrine than an incoherent one, ever since Locke, Hume, Darwin and many other post-Newtonian scientists concluded on the basis of the downfall of Cartesian physics that matter must now be conceded to be even capable of 'thinking' (cf. Yolton, 1983). The Cartesian notion of matter had excluded this, but it was now gone. With matter capable of thinking, materialism collapses (Chomsky, 2000). Locke's conclusion from Newton's post-Cartesian physics, that 'matter can think', was crucially *not* a materialist commitment but rather a statement to the effect that matter must in the end be conceded to contain all sorts of weird stuff, including 'thought'. As in the case of Newton's force of gravitation, we can't assume any more on *a priori* grounds that whatever naturalistic inquiry reveals will 'supervene' on (be entailed by) the mechanistic sort of matter that formed the basis for dualism in Descartes.[26] Gravitation, for one

---

[25] While one might identify structurally simple elements within distributed patterns of neural activity, single neurons say (Dehaene, 2002), their simplicity correlates unexplainedly with the simplicity of concepts and their intrinsic contents. Given the non-functionalizability of conceptual atoms, there is no faint hint on the scientific horizon of how neural correlates of *concepts* might be found, if these are to be more than 'brute' and unexplained correlates, and the concepts are not to stand merely in some or other causal relationship to their neural representations (such causal relationships few dualists including Descartes denied).

[26] For this reason, according to Chomsky, current materialism is in an embarrassing conflict with a widely endorsed (by scientists) conclusion in the eighteenth and nineteenth centuries, to the effect that a fully materialistic science had proved impossible (Lange, 1925; Yolton, 1983). Although that conclusion flies in the face of the received wisdom in the philosophy of mind (where the canon says that the reason for the downfall of dualism was the problem of psycho-physical interaction rather than the demolishing of the Cartesian notion of body through Newton's notion of matter), I think that this argument has to be respected, though it has not been widely discussed (see Hinzen, 2006a; Hinzen and Uriagereka, 2006).

thing, had been shown not so to supervene.[27]

But even if, with the new notion of matter, the basis for a coherent materialism was eroded, this did not and does not mean that within the new and encompassing notion of matter the question of how thought relates to non-thought is now answered. In fact, the matter is left as a mystery. And it is certainly a possible result of naturalistic inquiry that mental properties simply happen to be such that we do not expect a unification of mental and non-mental properties of this universe (see Conn Henry, 2005), given, in particular, the atomistic structure of the semantic, and the different structure of the non-semantic. An immediate reply to this would be that chemists until well into the twentieth century equally failed to expect a unification of chemical with non-chemical (physical) properties of this universe (Brock, 1992). But then, unification *did* occur, in the 1920s, through a conceptual revolution in the lower-level science, physics. Doesn't this show that just as it would have been pointless for chemists to conclude some sort of 'body–body' dualism in the nineteenth century, it is pointless for philosophers and linguists now to posit a 'mind–body' or 'semantic –non-semantic' dualism? Perhaps gravity, energy, electro-physiology and atomic particles as described by today's physics are not enough for unification, and other principles and primitives, now unknown, will have to be invoked to cope with psychology and the sciences of the human language faculty. But the answer is no, because the solution of the unification problem in the case of chemistry did not invoke reduction or elimination at all. Chemical theory set a constraint on what unification would have to achieve, and it did achieve that: chemical atoms were *there to stay*, though now under a physical description. Given my arguments against reduction, and the implausibilities of elimination, we should similarly expect that if a conceptual revolution in physics were to occur that enabled it to encompass psychology and linguistics, it would still be relevantly *true* to say that *there are* concepts, simple and complex, essentially semantic in the way we have claimed they are.

For now, then, our argument is that the 'mental' or 'semantic', as a particular domain of inquiry, is as 'real' as anything else is that naturalistic inquiry ever since the scientific revolution has brought to light: gravity, say, or chemical atoms. As far as we can tell now, there

---

[27] Modern relativistic physics, of course, has partially eliminated the weird 'action at a distance' again that led to the loss of the Cartesian notion of 'body'. But that was only for quantum mechanics to re-introduce it again. Today, still, the matter is controversially discussed, although it is certainly a possibility, and I think a mainstream view, that the necessary locality of physical interaction is *not* a property holding in this universe.

is, along with 'physical' particulate systems in nature, such as physical chemistry or genetics, a semantic one too, consisting of meanings as building blocks. If a future physics accommodates conceptual atoms much as it accommodated chemical ones in the 1920s, so be it: it is exceedingly unlikely that it will threaten the conclusion I call dualist here, namely that the realm of the conceptual is a domain of inquiry in its own right, irreducible to anything else, in a way that whatever results this inquiry yields on the basis of rational reflection and empirical evidence in linguistic theory, they should, for now, be interpreted *realistically*. At this point, that is, 'concept' should be regarded as a *theoretical primitive* with its own structural properties. Explaining concepts by something other than themselves (inferences, reference, representations, etc.), that is, seems misguided. Rather, a body of theory should be accumulated for concepts as primitive entities, much as chemists did for their atoms, ignoring their lack of a physical interpretation, and assuming that whatever physical interpretation *would* come up, would have to be constrained by the linguistic insights so far attained.

6. Relation to Qualia-based Arguments

The present argument proceeds on what we may call a *rationalist* path, in the sense that it takes as its basis a structural analysis of the organization of a *system of 'core knowledge'* (Gallistel, 1990; Spelke, 2000; Hauser & Spelke, 2004), the human linguistic system, as viewed in the biolinguistic tradition (Chomsky, 1959; Lenneberg, 1967; Jenkins, 2000). This point of departure contrasts sharply with that of the recent literature on consciousness and dualism, where we primarily find researchers introspecting their qualitative sensations — the 'what it is like' of their inner experiences. This latter approach has an *empiricist* flavour, though this time the relevant experience is an 'inner' rather than an 'outer' one. Could consciousness research benefit from a rationalist turn, in the same way that epistemology has with the arrival of the core knowledge hypothesis in psychology (see references cited above)?

   A rationalist (or Kantian) epistemology would note that without applying *concepts* — or *thinking* something — we are also not capable of *experiencing* anything (perhaps we are simply *processing information*): what experience a species has depends on what concepts it has. If so, experiencing qualia is *conditioned by* what is studied here, concepts. No concepts, no qualia, and no consciousness either. Rather than insisting on this Kantian epistemological

conclusion here, however, let us discuss whether qualia-based puz-
zles raised for the study of consciousness are perhaps *parasitic* on
the problem for concept possession discussed here.

   Consider Chalmers' (1996) claims on why life or phenomena like
digestion logically/conceptually supervene on the physical, whereas
qualia like the feeling of pain fail to do so. Logical or conceptual
supervenience means that it is not conceivable — it is a logical or
conceptual contradiction — that all the physical facts of the world
might be fixed, yet the non-physical ones in question (e.g., phenom-
enal or biological ones) could vary freely (i.e. not be fixed). Now, the
claims that life processes or pain do or do not logically supervene on
physical matter are *premised* by what our concepts of life and pain,
respectively, *mean*. It's because of what our concept of pain *means*
that we think a Zombie would not feel it. Perhaps, then, *concepts* are
the real agents behind the scene in familiar dualist arguments such as
Chalmers'. Consider Chalmers' *two-dimensional* framework in
semantics, which analyses concepts or meanings reductively as
kinds of *intensions*: these are mappings from possible worlds to
extensions (referents). One of these kind of intensions — the
so-called 'primary intensions' — specifies what a concept means or
refers to depending on which world *will turn out actual* (e.g.,
whether we will find ourselves on Twin Earth or Earth). Another one
— 'secondary intensions' — depends on which world *has turned out
actual*. E.g., it has been argued that after it *has* turned out after a
chemical analysis that we are on Twin Earth, not Earth, the term
'water' does not refer to the stuff found there, given that this stuff is
not H_2O but XYZ. In Chalmers' analysis, the latter (secondary)
intension presupposes the former (primary) one, for before we can
ask whether 'this watery stuff here' is XYZ or H_2O, we have to know
what stuff we are talking about: *watery* stuff (where the *primary*
intension is intended). That however presupposes our possession of
the concept associated with the primary intension, which in turn I
have been arguing is not specifiable without reference to the very
concepts whose possession is meant to be explained.[28]

   Nowhere in Chalmers are we being offered a reason for why these
concepts should be functionally analysable or reducible, let alone
those that are non-structured conceptual atoms. Chalmers (1996) does
not discuss these (unsurprisingly, perhaps, for functionalism, as a
metaphysical doctrine, if I am right, presupposes their absence, i.e. the

---

[28] See Bealer (2002) for another defence of the claim that the two-dimensionist reduction of
concepts has things backwards, depends on the concepts it claims to explain.

falsehood of the atomistic doctrine). But it even seems that one can construct a technical argument for the failure of logical supervenience of concepts that is formally analogous to Chalmers' argument against the logical supervenience of qualia, and to this I now turn. Consider our concept of water, whose 'primary' intension, with a meaning roughly like that of the word *watery stuff* (note the circularity of this specification). According to this primary intension, water *might* as well be XYZ, something we cannot exclude *a priori*: while most humans think water on Earth is H_2O, it is of course logically possible that there is no such thing as H_2O at all, even on Earth (it's possible that we have all been mistaken), and that the right formula is XYZ, or that there is *no* formula (say because water is no natural kind at all, like, probably, cancer, a real possibility). In other words, while there *can* be a dispute on whether something already identified as 'water' (using the primary intension) *really* deserves this name, given that it has been now found to be XYZ, water according to its *primary* intension basically *is* whatever *seems* to us as water. It is thus hard to see how, when studying water on Twin Earth, even the insight that we were dealing with XYZ there would take away the fact that it is the *watery* stuff on that planet that is XYZ. This indicates that we do *not* revise our primary intensions even in the light of unsuspected secondary ones. It seems we speak coherently (grammatically) when we say: 'Water on Twin-Earth is not H_2O'.

If, at the level of the primary intension, water basically is whatever *strikes us so* that we apply our concept of water to it, reference is circularly specified at that level, or by reference to the concept itself, as it must be, if I am right above, and *water* is a conceptual atom. For if it is an atom, then, although it will figure in various complex descriptions involving other atoms, it will not possibly *reduce* to any descriptions in which it doesn't itself figure. Its meaning would therefore *have*, exactly like that of all other atoms, to be 'rigid': all descriptive conditions (except the circular one of *being water*) seem to be essentially such that water in its primary meaning could either have *or* lack them, hence they cannot be *essential* to it (see Hinzen, 2006c, for this direct path from atomism to an explanation of rigidity, the former being what explains the latter). But this is exactly the conclusion that Chalmers (1996), with Kripke (1972), reaches for pain: because pain rigidly denotes what seems to us like pain (that is, a certain phenomenal experience), what seems to us as pain *is* pain. We cannot be taught by experience or experiment that we are mistaken in this, or that pain is something else, say a behaviour, or a neurological condition, to all of which pain would relate only contingently. Essentially it relates only

to itself, i.e. the experience it rigidly denotes. Logically or conceptually, that is, pain all and only supervenes *on pain*. Its reference can only be specified circularly. But now we see that in exactly the same way, water in its primary intension supervenes only on that intension. It is only a *contingent* fact what world we live in, and as long as we evaluate according to the primary intension, water *need not* be H_2O. It is only water, the substance, never water, the primary concept, that can *essentially* be H_2O. The primary concept can logically exist without the substance, just as the substance can logically exist without the concept (indeed, without any cognitive creature at all, since none might exist in a world with water). But, if a concept is logically/conceptually independent even of *having* an actual *physical* referent, let alone one with a particular chemical structure, *and* the existence of that referent is in turn logically/conceptually independent of their being any concept of it at all, we arrive at a failure of logical supervenience of ordinary human concepts at large.

In sum, it is entirely possible that the Chalmers-Kripke argument for dualism is, contrary to what Chalmers (1996) assumes, fully general and concerns concepts as such, quite independently of whatever inner phenomenology they are associated with. But what now fuels the argument for dualism is not phenomenology, but atomism. It is because concepts are non-functionalizable, and do not supervene logically either on their reference or on whatever external substances they correlate with or on their (non-semantic) mental representations, that dualism in my sense holds.

7. Concluding Remarks

Functionalists have assumed — crucially *with* the behaviourists they claimed to replace — that the *meanings* of the internal representations they posited could be somehow functionally or computationally explained (by causal roles or mechanical input–output relations). But (at least) the *primitives* of the combinatorial system of human language cannot be so explained, if indeed they are atoms. Contrary to widespread claims, generative grammar as evolving in the biolinguistic framework was crucially *not* committed to this functionalist stance on human concepts. In fact one can see the traces of essentially the argument from circularity that figures so centrally in the present paper in Chomsky's (1959) critique of Skinner's 'functional analysis' of linguistic behaviour. Perhaps this problem in any functionalist approach to human consciousness has never really been overcome.

When current NCC-approaches (e.g., Koch, 2004) aim to integrate semantics into their account of consciousness by connecting it to Churchland's associationist 'neurosemantics' (Churchland, 1998), their enterprise depends on assuming the falsehood of what is asserted here, the non-functionalizability of atoms and the logical independence of concepts on their physical representations or correlates. With many doubts on whether neurosemantics can account for the systematicity of human thought (Hadley, 2004) remaining, it is unclear why we should wed the NCC programme to the connectionist commitments of the neurosemantic programme that revive a behaviourist heritage (cf. comments on this heritage in Fodor and Lepore, 2002, chs. 8–9). After 50 years of functionalism, the relevant ontological choice *does* after all appear to be one between strictly eliminative materialism and a form of dualism. Functionalism as the putative intermediate option, which denies elimination while refusing to revise its physicalist ontology, may be a form of a metaphysical *mauvaise foi*.

To the extent and in the sense that the human conceptual system is linguistically conditioned and structured, *and* correlates with consciousness, my conclusion also supports Macphail's (1998) regarding the intrinsic language-dependence of our human kind of conceptual consciousness and thought. While a methodological naturalism will naturally start from a continuity assumption as regards human and non-human animal cognition, and much recent work in comparative psychology shows impressive and unsuspected cognitive feats on the side of animals (McGonigle and Chalmers, 2005), nothing in what we know rules out that human cognition (and linguistic cognition in particular) involves radically different principles, some of which may well correlate with consciousness in its human form.

References

Abler, W. (1989), 'On the particulate principle of self-diversifying systems', *Journal of Social and Biological Structures*, **12**, pp. 1–13.

Baker, M. (2001), *The Atoms of Language* (New York: Basic Books).

Barrett, J. (2006), 'A quantum-mechanical argument for mind–body dualism', forthcoming in Hinzen (2006b).

Bealer, G. (2002), 'Modal epistemology and the rationalist renaissance', in *Conceivability and Possibility*, ed. Gendler & Hawthorne (Oxford: Clarendon Press).

Block, N. (1995). 'The mind as the software of the brain', in *Thinking: An Invitation to Cognitive Science Vol. 3*, ed. D.N. Osherson and E.E. Smith (Cambridge, MA: MIT Press).

Borer, H. (2004/5), *Structuring Sense: An exoskeletal trilogy*; two volumes (Oxford University Press).

Bringsjord, S. & Zenzen, M. (2003), *Superminds, People Harness Hypercomputation, and More* (Berlin: Springer).

Brock, W.H. (1992), *History of Chemistry* (London: Fontana Press).

Carey, S. (2004), 'Bootstrapping and the origin of concepts', *Daedalus*, Winter, pp. 59–68.

Carey, S. (to appear). *The Origins of Concepts*. MIT Press.

Chalmers, D. (1996), *The Conscious Mind* (Oxford University Press).

Changeux, J.-P. (2004), *The Physiology of Truth* (Harvard University Press).

Chomsky, N. (1959), 'A Review of B.F. Skinner's *Verbal Behavior*', *Language*, **35** (1), pp. 26–58.

Chomsky, N. (2000), *New Horizons in the Study of Language and Mind* (Cambridge University Press).

Chomsky, N. (2005), 'On phases', ms., MIT.

Churchland, P.M. (1998), 'Conceptual similarity across sensory and neural diversity', *Journal of Philosophy*, **95** (1), pp. 5–32.

Conn Henry, R. (2005), 'The mental universe', *Nature* (July 7).

Dehaene, S. (2002), 'Single-neuron arithmetic', *Science*, **297**, 1652–3.

Dickinson, A., McGonigle, B. and Chalmers, M. (2003), 'Concurrent disjoint and reciprocal classification by *Cebus paella* in seriation tasks: evidence for hierarchical organization', *Animal Cognition*, **6**, pp. 185–97.

Fisher, R.A. (1930), *The Genetical Theory of Natural Selection* (Oxford University Press).

Fodor, J. (1998), *Concepts* (Oxford: Blackwell).

Fodor, J. (2000), *The Mind Doesn't Work That Way* (Cambridge, MA: MIT Press).

Fodor, J (2003), *Hume Variations* (Oxford: Clarendon Press).

Fodor (2004), 'Having concepts: A brief refutation of the twentieth century', *Mind and Language*, **19** (1), pp. 29–47.

Fodor, J. and Pylyshyn, Z. (1988), 'Connectionism and cognitive architecture', *Cognition*, **28**, pp. 3–71.

Fodor, J. and Lepore, E. (2002), *The Compositionality Papers* (Oxford University Press).

Fodor, J. and McLaughlin, B. (1995), 'Connectionism and the problem of systematicity: Why Smolensky's Solution doesn't work', in *Connectionism: Debates on Psychological Explanation*, ed. C. Macdonald. and G. Macdonald (Oxford: Blackwell).

Gallistel, C.R. (1990), *The Organization of Learning* (Cambridge, MA: MIT Press).

Gallistel, R. (1998), 'Brains as symbol processors: The case of insect navigation', in *Conceptual and Methodological foundations*. Vol. 4 of *An Invitation to Cognitive Science* (D. Osherson, series editor), ed. S. Sternberg and D. Starborough (Cambridge, MA: MIT Press).

Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A. and Trueswell, J.C. (2005), 'Hard words', *Language, Learning and Development*, **1**, pp. 23–64.

Hadley, R.F. (2004), 'On the proper treatment of systematicity', *Mind and Machines*, **14**, pp. 145–72.

Hauser, M.D. and Spelke, E. (2004), 'Evolutionary and developmental foundations of human knowledge: A case study of mathematics', in *The Cognitive Neurosciences, Vol. 3*, ed. M. Gazzaniga (Cambridge, MA: MIT Press).

Hinzen, W. (2006a), *Mind Design and Minimal Syntax* (Oxford University Press).

Hinzen, W. (ed., 2006b), *Prospects for Dualism: Interdisciplinary Perspectives*, special issue of *Erkenntnis* (forthcoming).

Hinzen, W. (2006c), *An Essay on Naming and Truth* (Oxford University Press, forthcoming).

Hinzen, W. and J. Uriagereka (2006), 'On the metaphysics of linguistics', in Hinzen (2006b).

Hume, D. (1739–40), *A Treatise on Human Nature* (Oxford: Clarendon 1978).

Jackendoff, R. (2002), *Foundations of Language* (Oxford University Press).

Jenkins, L. (2000), *Biolinguistics* (Cambridge University Press).

Koch, C. (2004), *The Quest for Consciousness* (Englewood, CO: Roberts & Co.).

Kripke, S. (1972), *Naming and Necessity* (Harvard University Press).

Lange F.A. (1925), *The History of Materialism* (London: Kegan-Paul).

Larson, R. and Segal, G. (1995), *Knowledge of Meaning* (Cambridge, MA: MIT Press).

Lenneberg, E. (1967), *Biological Foundations of Language* (New York: Wiley).

Leslie, A. (2000). 'How to acquire a "representational theory of mind"', in *Metarepresentations*, ed. D. Sperber (Oxford University Press).

Lewis, D. (1994), 'Reduction of mind', in *A Companion to the Philosophy of Mind*, ed. S. Guttenplan (Oxford: Blackwell) pp. 412–21.

Macphail, E. (1998), *The Evolution of Consciousness* (Oxford University Press).

McGonigle, B. and Chalmers, M. (in press), 'Executive functioning as a window on the evolution and development of cognitive systems', *International Journal of Comparative Psychology* (forthcoming).

Marcus, G. (2001), *The Algebraic Mind* (Cambridge, MA: MIT Press).

Meixner, U. (2004), *The Two Sides of Being* (Paderborn: Mentis).

Metzinger, T. (ed. 2000), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (Cambridge, MA: MIT Press).

Mufwene, S. (2001), *The Ecology of Language Evolution* (University of Chicago Press).

Papineau, D. (2004), *Thinking About Consciousness* (Oxford: Clarendon Press).

Penrose, R. (1994), *Shadows of the Mind* (Oxford University Press).

Rey, G. (2004), 'Fodor's ingratitude and change of heart?', *Mind & Language*, **19** (1), pp. 70–84.

Smolensky, P. (1995), 'Connectionism, constituency and the language of thought', in *Connectionism: Debates on Psychological Explanation*, ed. C. Macdonald. and G. Macdonald (Oxford: Blackwell).

Spelke, E.S. (2000), 'Core knowledge', *American Psychologist*, **55**, 1233–43.

Stapp, H. (2005), *A Mindful Universe*, Book MS., http://www-physics.lbl.gov/~stapp/stappfiles.html.

Terrace, H. (2005), 'Metacognition and the evolution of language', in *The Missing Link in Cognition: Origins of self-reflective consciousness*, ed. Terrace and Metcalfe (Oxford University Press).

Uriagereka, J. (forthcoming), *Anchors*.

Yolton, J.W. (1983), *Thinking Matter: Materialism in 18th century Britain* (University of Minnesota Press).