Atomism, pluralism, and conceptual content

Abstract: Conceptual atomists argue that most of our concepts are primitive. I take up three arguments that have been thought to support atomism and show that they are inconclusive. The evidence that allegedly backs atomism is equally compatible with a localist position on which concepts are structured representations with complex semantic content. I lay out such a localist position and argue that the appropriate position for a non-atomist to adopt is a pluralist view of conceptual structure. I show several ways in which conceptual pluralism provides an advantage in satisfying the empirical and philosophical demands on a theory of conceptual structure and content.

1. Introduction

What sort of structure do our concepts have? A first pass at an answer suggests two broad possibilities: *some* kind of structure, or *no* structure. Many philosophers, psychologists, and linguists suppose that the first reply is correct. In particular, many of them suppose that concepts possess semantically relevant constituent structure: they have parts that contribute in some way to their content. Recent debates over concepts have mainly centered around what sort of constituents they have. Arguing against this consensus are advocates of the second reply. These conceptual atomists hold that concepts lack constituents. They present a host of arguments intended to show that once we consider the range of constraints a theory of concepts and mental content must satisfy, atomism is the only game in town.

The division between atomistic and non-atomistic theories of concepts is a fundamental one. I will argue here that at least some arguments taken to cut decisively in favor of atomism fail to do so. Atomism is fundamentally a defensive position: it is motivated by the purported inadequacies of the alternatives.[1] I consider three arguments in particular: the principled basis argument, the argument from what I call the 'subtraction strategy', and the argument from

[1] This dialectical fact is displayed especially clearly in Fodor (1994).

analyticity. My goal is to show what set of commitments a non-atomistic theory should adopt to avoid the atomist's objections. I'll argue that, both in terms of explaining the psychological data concerning concepts, and in terms of answering these atomistic arguments, a theory of concepts should be semantically *localist* and *pluralistic*. I lay out some properties of this pluralistic localism and defend it against objections. The conclusion is that atomism has no advantages over a suitably modified localist theory of concepts.

2. An outline of atomism

Following standard practice, I will mainly discuss *lexical* concepts here.[2] Although concepts are expressed by lexical items, I do not assume that this is one of their essential features. Creatures lacking a language, or people possessing a different language, can share concepts with English-speakers. Finally, although concepts are expressed by lexical items, I don't assume that they are word meanings. The constraints on a theory of meaning for natural language might be different from those on a theory of concepts, even if language expresses concepts.

Atomism is the following thesis about lexical concepts:

(Atomism) A concept's identity is not even partially constituted by its relations to any other concepts.

These relations can be either mereological or inferential. Taking mereological relations first, atomism is inconsistent with C_1's being individuated by any of: (i) C_1 having C_2 as a part; (ii) C_1 being a part of C_2; and (iii) C_1 and C_2 being co-hosted by C_3 (where neither C_1 nor C_2 is a part of

[2] Saying perspicuously what lexical concepts are proves to be tricky. Some concepts expressed by single words are arguably complex. 'Pools' is a single word containing the morphemes 'pool' and '-s', hence it may consist of the concepts POOL and PLURAL. 'Mailman' is a single word, but a compound of 'mail' and 'man', expressing the concept MAN WHO DELIVERS THE MAIL. I will largely stick to discussing concepts expressed by free monomorphemic expressions. A note on notation: small caps are used to form the names of concepts, angle brackets ('<', '>') are used to form the names of contents.

the other). If COW had ANIMAL as a constituent, then its identity would be at least partially constituted by its relation to that constituent. By the same token, atomism entails that a concept maintains its identity no matter what concepts it enters into combination with: COW remains the same concept when it occurs in BROWN COW and DANGEROUS COW. Finally, COW and HERBIVORE retain their identity whether or not they are co-hosted in the representation COWS ARE HERBIVORES. So one could have COW without having HERBIVORE or linking COW with it.

A second kind of relation is being connected by an inferential disposition, e.g., the disposition to infer that things falling under one concept also fall under another. WATCH and BATTERY are related inferentially if I am disposed under some circumstances to infer from *X* IS A WATCH to *X* HAS A BATTERY (where this may be an inference mediated by other beliefs). Whatever inferential relations WATCH bears to other concepts, they do not affect its identity as that very concept. Atomism allows that WATCH can play an almost arbitrarily bizarre role in one's beliefs and inferences.

Three other theses are frequently associated with atomistic theories of concepts:

(Ref) Conceptual content is just reference.

(Indiv) Concepts are individuated solely by their content and their formal/syntactic properties.

(Inf Sem) Conceptual content is determined solely by information-carrying causal relations between concepts and their referents.[3]

[3] Inf Sem needs to be restricted at least to the non-logical vocabulary. Even atomists tend to favor inferential or causal role semantics for logical concepts. Indeed, the one class of concepts for which inferential role semantics is thought to be plausible is logical (and possibly mathematical) concepts. I will restrict my attention here to non-logical concepts, putting aside the possibility that atomists must have adopt a disjunctive theory to deal with the full range of human concepts.

These claims can be found in the works of several prominent atomistic theorists (Fodor, 1987, 1989, 1990; Margolis, 1998a; Millikan, 1998, 2000).[4] The thread that ties these theses together has to do with the individuation and possession conditions for concepts. Ref plus Indiv entail that concepts are individuated by only reference and syntax.[5] Having a concept is just a matter of having a psychological primitive with a certain formal/syntactic profile and content (reference), where the latter depends only on a structure's nomic referent-to-concept causal relations, not on the details of how these relations are mediated, including what other concepts happen to be involved in the concept's activation.

3. Localism and cognitive models

One of the central arguments against allowing concepts to be individuated in terms of relations to other concepts invokes the threat of holism. This Principled Basis Argument is familiar.[6] Suppose that conceptual identity (or some aspect thereof, e.g., conceptual content) is determined by a concept's relations to other concepts. Then either relations to only a subset of other concepts are relevant to determining identity, or relations to all other concepts are relevant. But there is no principled way to delimit only a subset of a concept's relations as being identity-determining. Having included some of these relations, we have no non-*ad hoc* reasons to exclude others. So *all* of a concept's relations must be identity-determining, and we have holism: the doctrine that all of a concept's relations to other concepts constitute its identity conditions. But

[4] However, not all atomists endorse all of these claims. Millikan, for one, wants to resist the idea that a mental representation's formal properties play a role in its individuation conditions. Also, her own psychosemantics incorporates a teleological component as well as an informational component. I will generally ignore such disagreements among theorists who believe enough of the core bundle of atomistic theses.

[5] Aydede (1998) argues that, in fact, there is a dilemma about what fixes a concept's syntactic properties. Either these properties are physically determined, in which case a strong type-physicalism follows, or they are relationally determined, in which case a form of holism—about syntax, not about content—follows. Both of these are undesirable from Fodor's perspective. I will discuss the dispensability of syntactic properties for certain individuative purposes in Section 6.

[6] See Fodor (1987, Ch. 3), Fodor & Lepore (2002) for discussion. Block (1993) refers to this argument as 'the Old Hat'.

holism is undesirable, given that it blocks sharing of concepts across different cognitive economies (different persons, different temporal stages of a single person), and hence blocks us from formulating widely applicable psychological generalizations.

Localism holds that some but not all of a concept's relations to other concepts are constitutive. A localist reply to the principled basis argument comes from considering the kind of explanatory models offered in psychology to account for concept-involving phenomena such as reasoning and categorization. The salient fact here is that the kinds of models of conceptual capacities that have a significant degree of empirical support typically treat concepts as having local semantic structure. This fact forms the basis of the following Argument for Localism:[7]

1. A range of well-supported psychological models posit conceptual representations that have complex structure.
2. One aspect of this structure is plausibly interpreted as local semantic structure.
3. In general, when we have a model that is capable of accounting for some range of psychological phenomena, we should take it to be at least a provisional guide to the structure of the psychological mechanisms that underlie the production of those phenomena.
4. *Conclusion*: We have at least provisional reasons to suppose that concepts are locally semantically structured.

In the remainder of this section I will sketch some assumptions common to many of these models, with one caveat: I won't discuss the third premise in any detail. The guiding assumption here is that cognitive models that enjoy a significant degree of empirical validation are *prima*

[7] The suggestion that the cognitive psychology of concepts could provide a response to the Principled Basis Argument is also put forward by Pereboom (1995). I agree with his emphasis on allowing the relevant sciences to discover the microstructure and behavior of the kinds that they cover (the science in this case being psychology, the kinds in this case being concepts and thoughts).

facie likely to be accurate descriptions of the psychological states, processes, and mechanisms underlying the task in question. This premise or something like it is taken for granted by all non-instrumentalist ways of understanding cognitive models, so I assume that it will be relatively uncontroversial in this context.

The first premise is supported by looking at the structure of particular empirical models of concepts. A cognitive model describes (i) a set of representations, (ii) processes that operate over those representations to effect performance on a task, and (iii) resources that are available to assist in processing (e.g., limited capacity memory and attention). Behaviors are explained, in part, by the operation of these processes, drawing on these resources, over these representations. Cognitive models of concepts posit that when subjects carry out tasks involving, for example, recognizing instances or drawing inferences about a category, the category representations they use are *complex*. In lieu of a thorough review of the literature, several paradigmatic examples can be cited.[8]

Traditional prototype models provide one case. On these models, concepts contain features that encode information about the statistically central properties of the category represented.[9] Prototypes have been widely implicated in our performance on a variety of cognitive tasks, including perceptual classification, lexical inference, judging typicality, category learning, and inductive inference. The explanation for subjects' performance appeals to a variety

[8] See Murphy (2002) for a masterful survey of empirical studies, models, and methods pertaining to the psychology of concepts. Markman (1999) also describes in detail a variety of particular approaches to cognitive modeling of concepts, memory, language, and perception. One note on the kinds of models being considered here: I am restricting my attention to models that are fairly 'traditional': that is, classical symbolic models, rather than connectionist or dynamical systems models. I think that many of the claims about semantic structure that are made here could equally be made about connectionist models of concepts and categorization. However, recasting all of the arguments again in the appropriate form would be time-consuming. In addition, it might obscure important issues. Connectionist models raise special problems of their own, which deserve separate discussion. So in the interest of keeping the discussion manageable, I focus on those models that employ roughly classical symbolic representations.

[9] This need not be the *actual* central tendency of the category of course, but only the central tendency from the subject's epistemic perspective.

of similarity-based computations over prototypes: for instance, categorizing something under TAPIR involves computing its similarity (in some respects or other) to the tapir prototype.

Another example involves research on naive causal models (Rehder, 2003a, 2003b; Sloman, 2005). Causal models are structured representations consisting of a set of features plus a set of dependency relations among those features. These dependencies represent how the subject takes properties in the domain to be related causally to each other (e.g., in cats, happiness causes purring; hunger causes scratching, feeding causes happiness, etc.). In causal models proper the relations are all causal, but generalizations of such models recognize logical dependencies as well as unspecified generic dependency relations. Causal models can in some instances predict performance better than prototype-based models do. Causal models involve categorization by shared fundamental causal properties, rather than just similarity computation. There is good evidence that people both possess causal models for many commonly cognized domains, and that they can readily be trained to generate models of novel domains. The point about these models, for present purposes, is that they are relatively circumscribed representations that explain subjects' performance in a range of tasks, including typicality judgments, imagining counterfactual instances, categorization, and induction (Sloman, Love, & Ahn, 1998). Most importantly, they can be individuated in local terms, on the basis of their constituents and the relations among them. Not every piece of information that one possesses about a category is part of the causal model of that category that one uses in a particular task.

Finally, it has also been proposed that concepts, particularly concepts of natural kinds, represent essential properties of those categories (Gelman, 2003). According to this psychological essentialism, concepts represent two kinds of information about a category: superficial characteristic properties and essential ones (Medin & Ortony, 1989). Characteristic

properties include things like appearance and typical behavior; essences include whatever unperceived and perhaps unknown properties cause the presence of the superficial properties, like having certain insides or certain genetic structure (Gelman & Wellman, 1991). Substantial evidence suggests that even very young children tend to view certain domains (e.g., living things) in terms of essences, and even that they have specific ideas about what these essential properties are (although these ideas may be slow in developing; see Simons & Keil, 1995). Essence representations can be seen as causal models that have a specific form: the superficial characteristics are caused by the essential properties, which in turn are necessary and sufficient for category membership. Superficial characteristics are neither necessary nor sufficient. Again, essentialist representations are like other causal models in being individuated in local terms.

These cases show that many standard models of concepts in psychology treat them as complex but limited-capacity representations. Barsalou (1992) proposes that the representations posited in such models can generally be thought of as *frames*.[10] Frames are structures containing (a) pairs of attributes and values (e.g., COLOR: RED, TEXTURE: ROUGH), and (b) relations among those pairs. These are properties of how frames are organized: they go beyond simple unstructured lists of features. Relations can include causal and other dependency relations among features; so, for instance, WHEEL might contain not only ROUND and CAN ROLL, but also a link indicating that roundness causally supports the ability to roll. Frame representations are extremely general and powerful, and have application across a variety of different kinds of concepts, including, e.g., action concepts (Fillmore, 1968) and ad hoc concepts (Barsalou, 1983).

The existence of a variety of empirically validated frame-based cognitive models provides a reply to the principled basis worry. The best cognitive models of conceptual

[10] For an extensive discussion of the properties of frames and the empirical evidence in favor of feature-and-frame based representations of concepts, see Markman (1999).

processing are, and ought to be, weighty arbiters of what sort of structure concepts have. This is as it should be if the structure of concepts is an empirically decidable question at all. The principled basis that these theories deliver is that concepts are individuated by relation to the relatively small set of features that are their constituents. What those features are is determined on the basis of the particular cognitive model being considered: in an exemplar model, the features are those that characterize representations of particular category instances, in an ideal-based model (Burnett, Medin, Ross, & Blok, 2005), they are those that characterize a normatively ideal category member, and so on. Many different kinds of information can potentially be incorporated into our concepts. But each of these concepts is a limited bundle of such information, functioning in a way characterized by a particular cognitive model.

There is another version of the principled basis worry that might be pressed. One might argue that even if concepts are individuated on the basis of (*inter alia*) their constituent structure, that leaves open the question of why some features rather than others are constituents of a particular concept. What is the basis for one feature's being included in a concept rather than another? This question differs slightly from the previous one. It is one thing to ask about the principled basis for restricting concepts to limited bodies of information; it is another to ask about what makes a particular representation a part of a particular concept.

The answer to the second question must be that the inclusion or exclusion of particular features is just a matter of how the conceptual system is geared towards acquiring, retrieving, modifying, and constructing representations of categories. The particular rationale for a set of features being packed into a concept might include the fact that they are frequently used in reasoning about the category, or goal-driven considerations (i.e., the fact that they are especially relevant to the particular task for which they are being tokened). Perhaps past success in

representing a category in a certain way makes it more likely that the information used in these past episodes of thought will become part of the concept of that category in future episodes. Other explanations are possible, but all of them make reference to psychological factors to delimit concept-constitutive from non-constitutive relations. There is no need to look for a deeper answer to the principled basis argument than the one provided by theories about the mechanisms of concept formation and deployment.

Atomists have protested that the psychological structures described in the sorts of cognitive models I am discussing, while real, cannot really be *concepts*. The reason is that concepts are the building blocks of a system of psychological states that are productive and systematic (Fodor & Lepore, 1996). These properties of the attitudes are best explained by the compositionality of concepts. Compositionality has many possible formulations, but the basic idea is that complex concepts can be constructed out of simpler ones via a set of recursive construction rules that have access only to the simple concepts themselves. So it is sufficient to possess a complex concept that one possesses the simple concepts plus the combination rules. By applying the rules to the simples, one can produce indefinitely many complexes by a simple mechanical procedure. There should be no constituents in the resultant complex that are not constituents of the simples. The compositionality objection to treating concepts as complex is that these structures do not combine compositionally. The standard examples involve prototypes, but can be generalized: the PET FISH prototype contains so-called emergent features, such as EATS FISH FLAKES and LIVES IN A BOWL, not found in the PET or FISH prototypes; other examples can be generated freely. But combining compositionally is a necessary condition on something's being a concept. And so even if these structures are useful in explaining categorization, reasoning,

analogy, and so on, they cannot genuinely be concepts, even if they are part of the machinery by which concepts are actively deployed.

An enormous amount has been written recently concerning compositionality arguments, and I cannot discuss the argument in great detail here. However, I believe that the broadly correct response has been given by Prinz (2002, Ch. 11) and Robbins (2002). In essence, both point out that compositionality is a modal property: it concerns whether concepts *can* combine in a certain (non-emergent) way, not whether they always do. The fact that many complex phrasal concepts contain emergent features does not cast doubt on whether they can do this. Even if all phrasal concepts *de facto* contained emergent features, this would not show that they couldn't combine compositionally. Frame representations, moreover, are tailor made to be able to combine recursively into larger structures. Frames can contain other frames as constituents, and frames can fill roles in propositional schemata. There is every reason, then, to think that they are capable of purely compositional combination, even if this non-emergent form of combination is *de facto* rarely used. Compositionality and concept combination more generally deserve a more thorough discussion than this, but this will have to suffice for the moment.

4. Referential and cognitive content

A further question concerns how frame representations should be semantically interpreted. This issue is not generally addressed in the psychological literature, and it goes to the question of what support there is for the second premise of the Argument for Localism. So far I've argued that in well-confirmed cognitive models, concepts are individuated in terms of local relations; it remains to be seen that these relations support semantic structure. In the first place, the conventional labels assigned to concepts in cognitive models reflect their representational content. A frame is a TELEPHONE or MOLE concept because it represents telephones or moles.

This is just to say that frames in cognitive models are representations. Moreover, though, frame models describe concepts as containing features. These features are themselves representations of categories. Being representations, they are individuated in part by their content. So concepts can in many circumstances be usefully modeled as frame representations that are individuated by at least two factors: (a) what they represent; and (b) the features they contain and how these are organized. This much is presupposed in psychologists' descriptions of these structures.

But if conceptual features themselves have content, and this content partially individuates them, then a consequence of (b) is that conceptual representations are *also* partially individuated by their features' content. A natural suggestion is that these content-bearing constituents contribute their content to their host representations in some fashion. The features a concept contains may be seen as encoding the *way* that the concept represents what it does. These features express the properties that the thinker grasps in representing the class of things that the concept as a whole picks out. Existing psychological models of concepts, then, are plausibly interpreted as ascribing to them two semantic contents, one of which we can call the concept's *referential* content (RC), the other of which is a reflection of the way that the concept's constituent features are configured. Call this latter content *cognitive* content (CC). Tokening a concept involves representing its referent under the properties given by its cognitive content.

On this view, it is appropriate to say that each concept has *two* sets of truth conditions. Contents are properties or other aspects of the world that stand in semantic relations to concepts. Concepts refer to, and hence are true of, their extensions (their R-contents). They are also, though, true of whatever things instantiate the properties picked out by their cognitive contents. In entertaining a complex concept C we are representing Fs (C's R-content) as being Gs (C's C-content). The formula 'Subject *S* represents *F*s as being *G*s at time *t*' contains two open place-

holders for identifiers of the things represented by the concept's two truth conditions. It thus gives us a natural way to state these dual truth conditions.

Dual truth conditions also give two dimensions along which a particular application of a concept to an object *x* can be assessed for veridicality. Given two representational contents, one might erroneously represent an object *x* (a non-*H*) as being *H*, where *H* is either the concept's R-content or C-content. So one could perceive a ferret and think falsely that it is a cat in virtue of its satisfying a weighted majority of the cognitive content of one's cat concept. This would be a misapplication at the level of R-content. Alternately, one could perceive a cat and think truly that it is a cat, for the same reason that one misclassified the ferret. But if the cat turns out to be tailless, this act might still involve an element of misrepresentation, since if one's cat concept contains the feature HAS A TAIL one would be representing *this* cat as having a tail. This would be a misapplication at the level of C-content. As an analogy, consider the representational qualities of drawings. A drawing may resemble Walt but not be a drawing of Walt (the first case), or be a drawing of Walt but not resemble Walt (the second case). That we recognize these two dimensions of possible error—errors in what is represented and errors in how it is represented—suggests that the RC/CC distinction tracks some of our pretheoretic distinctions concerning content.

Are there any constraints on how these two contents may be related? Five possibilities exist for any concept (assuming RC and CC $\neq \varnothing$):

1. RC = CC;
2. RC $\subset$ CC;
3. CC $\subset$ RC;

4. $RC \cap CC \neq \varnothing$ & $\sim[(RC \subset CC) \vee (CC \subset RC)]$;

5. $RC \cap CC = \varnothing$.

Call a concept *adequate* iff its cognitive content and its referential content have a nonempty intersection but don't coincide completely (cases 2-4), and *fully adequate* iff its cognitive content and referential content coincide completely (case 1). If Sam's SQUARE concept has <equilateral rectangle> as its cognitive content, it is fully adequate, since all and only squares are equilateral rectangles. It is clear that simple descriptivism, on which referential content just *is* whatever category is picked out by cognitive content, won't do in general. This is illustrated by is the familiar and widely accepted facts of ignorance and error (Kripke, 1980; Putnam, 1975).

Adequate concepts can be overextensions, underextensions, or intersections. If Sam's OTTOMAN concept has as its cognitive content <kind of furniture>, it is adequate despite the fact that this content also applies to sofas and dressers, since ottomans are a kind of furniture (case 2). Overly expansive concepts can be perfectly useful. An EMERALD concept that represents emeralds as <green transparent precious gemstones> doesn't distinguish emeralds from other gems such as green garnet and green tourmaline. That doesn't diminishes its utility in reasoning about and identifying emeralds—perhaps the most frequently encountered objects in the environment that satisfy that cognitive content happen to be emeralds—even if it clearly fails to qualify its possessor as a gemologist. Underextended concepts (case 3) can be useful as well, and for similar reasons. Consider a child whose FORK concept has the cognitive content <tined utensil in my house> (this example adapted from Higginbotham, 1998). If all forks are tined utensils, this concept's C-content applies only to *local* forks, which may be enough for the children to manage the fork encounters she is likely to have. Intersective concepts (case 4) are exemplified by prototypes, which famously pick out some common members of their R-content, but also

superficially similar members of other categories (e.g., typical zebras as well as cunningly painted horses). Cognitive contents may only pick out superficial or merely locally useful properties of a category, rather than fundamental individuating information. This is unsurprising, given that concepts are primarily ways of representing the world that are geared towards the satisfaction of our practical interests and goals. Given the cognitive effort involved in discovering more genuinely individuating information about categories, it is sensible for concepts to encode the least effortful information possible.

Finally, concepts may be *inadequate*: their cognitive contents may not include any of their extension (case 5). Suppose Sam conceives of cats under the cognitive content <furry meowing animals kept as pets>. If Sam lives in a world where cats themselves don't meow, but are host to a special kind of symbiote that produces that distinctive noise when the cats open their mouths, then Sam's concept refers to cats despite being inadequate. Further, inadequacy need not result in practical harm, for instance if it never matters to any of Sam's cat interactions that his cat-concept is wrong about the source of meows. Referential content ought to be robust enough to tolerate some misinformation and slippage in our concepts, otherwise we could refer only to what we had fully adequate information about. This is one aspect that distinguishes this view of conceptual content from a Fregean view: cognitive content does not determine reference.

Since none of these five cases can be ruled out as possibilities, the RC/CC relation is a loose one. Only case 5 presents a potential difficulty, since as stated it allows for potentially bizarre mismatches of RC and CC, such as a concept that refers to rhinoceros but has as its cognitive content <purple & fuzzy & round & grows on trees>. Even the occurrence of these cases, however, could potentially be accounted for by mechanisms of deference. If I know nothing of rhinos and a trusted source says to me 'Rhinoceros are purple, fuzzy, round, and grow

on trees', I may form a concept that refers to rhinoceros (through a mechanism of linguistic deference or referential borrowing) but which represents rhinoceros by a strange cluster of properties that no rhino possesses. I say I *may* form such a concept because whether it is a genuine possibility depends on what fixes reference. This issue is unsettled, to say the least. Language provides a possible mechanism for the acquisition of bizarre concepts like this in virtue of its power to misrepresent: to convey reference to combinations of properties without regard to whether they genuinely co-occur in the category being picked out. But whatever we say about bizarre cases, we are accustomed to accepting less outlandish misrepresentations in many everyday situations. If everyday attributive practices are a guide, there appear to be relatively few constraints on how cognitive and referential content can possibly be related (essentially this point is made by several of the examples in Burge, 1979).

Despite this looseness, two norms address the appropriate relation between R-content and C-content. These are the norm of accurate representation and the norm of efficient representation. Accurate representation prescribes that C-contents generally ought to aim at adequacy, with full adequacy being the ideal. This norm derives from the fact that you may be most successful in deploying your F-concept if you represent Fs under properties that they actually have. The most reliable way to detect Fs, for instance, is to have in mind those properties that perfectly distinguish Fs from all other categories. Efficient representation, on the other hand, prescribes that C-contents should be manageably deployable given the demands of memory and processing ability, and the cost of one's epistemic labor in constructing them. Efficiency stems from the fact that working memory is a limited resource, and from the fact that fine-tuning one's representations to achieve better task performance involves dedicating extra time and resources that might be better spent elsewhere. Resource constraints like the size of

working memory tend to promote representations that are smaller and easier to deploy, while constraints of epistemic labor (e.g., how hard it is to achieve full adequacy) tend to permit a certain amount of inaccuracy in one's representations, where correcting these inaccuracies would be highly costly.

Accuracy and efficiency can potentially conflict with one another. Where accuracy pushes towards full adequacy, efficiency will trade off towards adequacy, if adequate representations can do the relevant job. Hence it wouldn't be surprising to find that most of our concepts fall in the range of mere adequacy. Even inadequate concepts, as noted, may be tolerated if the inadequacy is never noticed or never matters to task performance. If these norms are actually respected by the conceptual system, we would expect to find that, depending on how important they are in a creature's cognitive and practical endeavors, concepts exhibit varying degrees of fit between RC and CC. Whether, or to what extent, these norms govern how our concepts work is, of course, an open question.

Cognitive content is closely tied to the explanation of reasoning and behavior, since cognitive content derives from the features of a concept, and in cognitive models it is these features that partially fix a representation's causal role. We should note, though, that featural structure only determines *part* of a concept's causal role in conjunction with processes that can access these features and resources that constrain the operation of these processes (Anderson, 1978). So having a certain causal role is a *derivative* property of a concept, not a fundamentally individuating property. Causal role derives from the intrinsic structure of a concept in combination with its cognitive surroundings. This means that structured concepts are portable, in that they may be embedded in different belief systems and different inferential dispositions and systems of psychological processes without changing their identity. A cognitive model defines a

set of representations, processes, and resources, but the representations defined therein may also be employed in other models with different processes.

Cognitive content should not hastily be identified with *narrow* content. A concept's features represent an organized set of properties that constitute the concept's cognitive content. But it is perfectly imaginable that these features themselves might have *wide* representational content. Cognitive content is a semantic way of characterizing how a person represents an aspect of the world. This emphasis on the person's epistemic perspective may suggest that it should be interpreted narrowly. There are two reasons to resist this. First, it has been argued persuasively by White (1982), Lepore and Loewer (1986), and Taylor (1993) that there are *no* properties that cannot be 'twinned'; consider White's twin-liquid case, or Lepore and Loewer's twin-red case. Given the difficulties faced by the notion of narrow content, leaving open the possibility that all conceptual content is wide seems advisable.[11]

Second, advocates of the present theory have independent reason to think all contents are wide. I have been arguing that lexical concepts are complex, and that in virtue of this complexity they have two contents that serve different functions and may diverge. Concepts are capable of being combined into larger structures, e.g., phrasal concepts and full propositional judgments.

[11] In a series of papers, Chalmers has developed a sophisticated two-dimensional semantics for thoughts (Chalmers, 2002a, 2002b, 2003). He argues that content decomposes into epistemic and subjunctive intensions, and shows that these can play a role in addressing a number of traditional problems concerning thoughts and their attribution. For two reasons, cognitive content as elucidated here should not be identified with Chalmers' epistemic intensions. First, epistemic intensions are composed solely of 'semantically neutral' descriptions of situations, where a semantically neutral description is one that is immune to twinning. But it's unclear, first, whether there really are any such semantically neutral descriptions; and, second, even if they exist, it is doubtful whether they can be couched in a conceptual vocabulary that any of us actually possesses. Cognitive contents, by contrast, must be made of concepts available to the thinker. Second, epistemic intensions are determined by the thinker's idealized a priori rational reflection on epistemically available possibilities. This idealized rational reflection is, sadly, a far cry from the deliberative powers that most of us possess. What isn't clear, then, is how an epistemic intension can make contact with the everyday, non-idealized behavior of thinkers if it is determined only by such an idealized reasoning procedure. This seems like a problem if a thought's epistemic intension is central to explaining actions. None of this is to argue that there aren't epistemic and subjunctive intensions of thoughts, of course, or that they can't do some of the things Chalmers claims they can. It's just to illustrate some ways in which cognitive contents and epistemic intensions serve different theoretical functions.

When concepts combine, they contribute both their cognitive and referential contents. So CATS CHASE MICE is built up from CATS, CHASE, and MICE. Its R-content is the proposition <cats chase mice>, and its C-content is the complex proposition built up from the set of properties contributed by each constituent, that is, the proposition that expresses the way in which the subject represents the situation of cats chasing mice. The familiar semantic rules of composition in each case produce, in parallel, complex contents from the contents of the constituents. R-contents of concepts combine to form R-contents of judgments, and similarly for C-contents, although neither C-contents nor R-contents determine the other.

But not all concepts can be complex; there must be some primitives. What are we to say about their content? For complexes, C-content and R-content may diverge. For primitives, though, we should say that they coincide. So, just to pick an example, if CAUSE turns out to be primitive, then both its C- and R-contents are <cause>. If, *contra* what I've argued here, CAT is primitive, then both of its contents are <cat>. The justification for this assignment comes again from considering the fact that C-content encodes how concepts represent their R-contents. It is uncontroversial that CAT refers to cats. If a CAT primitive represents them in any way at all, presumably it is just *as such*, that is to say, as cats. The alternative is to say that primitives represent their R-contents in *no* way; that is, their C-content is empty. But since C-content for complexes is built up from that of primitives, if the primitives lacked C-content, no complex could have it either. Given the bottom-up compositionality of C-content, this option is undesirable. Some C-content is needed at the bottom level for there to be C-content at any higher level.

Now, assuming that R-content is a form of wide content, as reference is usually taken to be, the argument for all content being wide runs as follows. For primitives, C-content is R-

content. All concepts can be decomposed, ultimately, into primitive representations in a certain sort of combination. So the C-content of a complex concept is a function of the C-content of its constituent features. At the bottom level, this C-content is identical to referential content, hence is a form of wide content. Assuming that merely combining wide contents does not affect their status *as* wide, the C-content of *non*-primitives (e.g., lexical concepts) is wide content as well. So theory-internal bottom-up considerations about how to assign content to whatever conceptual primitives there are lead to the conclusion that cognitive content is a form of wide content. Cognitive content turns out to be referential as well on this picture. This also comports with the way that content is assigned to features in many typical psychological experiments. If participants are presented with pictures of imaginary creatures belonging to categories defined by the presence or absence of properties such as having a curly neck, having a spotted body, etc. (see, e.g., the materials in Brooks, 1987), the features attributed to their concepts by the modelers are determined by the properties to which they respond in the stimuli.[12]

An atomist may object at this point that localism actually *depends* on the truth of atomism. For localists, concepts depend for their identity ultimately on their R-content and their constituents, which fix C-content as discussed above. No external relations to other concepts are allowed to contribute to their type-individuation, on pain of re-introducing holism. But this

[12] See Prinz (2000) for an elaborate theory of content fixation on which concepts possess two wide contents. The view developed here, while it bears some similarity to Prinz's view, differs from it in several aspects. Most obviously, Prinz is committed to a form of concept empiricism on which all concepts decompose into sensory representations, and hence all cognitive contents are ultimately composed of properties represented by a creature's sensory systems (Prinz, 2002). The view developed here is not empiricist. In part, Prinz's empiricism is motivated by his advocacy of Inf Sem: content is fixed by nomic relations, which involve the use of concepts as structured detectors of distal properties. This seems to hamstring the theory of concepts by making central their role in detection over their more lofty roles in inductive inference, counterfactual and causal reasoning, organization of memory, and other offline deployments. Giving perceptual categorization pride of place in delimiting concepts' cognitive role seems to me mistaken. Another difference between C-content and Prinz's 'notional' contents (later also called 'cognitive') is that, for Prinz, notional contents are appearance properties: open disjunctions of properties that can be detected by a creature's basic perceptual apparatus. But this proposal doesn't jibe well with the suggestion that we ascribe notional contents in explaining each other's behavior, or that psychologists characterize participants' attitudes by reference to such contents, since our explanations do not seem to be disjunctive in the way that notional contents are.

means at least that the primitives that compose a concept must have semantically atomistic individuation conditions: their content cannot depend on internal relations to other concepts, since they lack constituents, and cannot depend on external relations, either. An identical argument applies to the determination of the R-content of lexical concepts, since it is not typically determined by C-content and (for the same reasons) not determined by external relations. So sufficiently portable localist concepts depend on the existence of an account of reference fixation that makes it largely independent of relations to other particular concepts. Causal or nomic relations, along the lines of Inf Sem, would seem to be the natural alternative candidate for fixing the content both of primitives and of complex concepts.

This line of argument, though, doesn't threaten semantic localism for lexical concepts. Semantic atomism is a claim about content or reference fixation. If R-content is always determined by a concept's causal relations with its referent, in abstraction from the details of how those relations are mediated, then R-content is atomistic. This doesn't establish the full doctrine of *conceptual* atomism, however, since it is compatible with semantic atomism about R-content that conceptual vehicles should be structured, and should possess other aspects of content (C-content) that are not determined atomistically, at least not in complex concepts. As noted above, the relationship between RC and CC is a loose one, just as would be predicted if this form of semantic atomism were true. The very considerations that were used to motivate this looseness of fit above are the ones that motivate atomistic theories.

5. The subtraction strategy and pluralism

I will now turn to objections to localism. One anti-localist stratagem involves appeal to everyday intuitions about concept possession. These intuitions are elicited as follows. Take any concept, say CANCER, that the localist claims is semantically structured. Suppose having the

concept CANCER is having a causal model of the disease. Now consider a person, Ada, who doesn't have that causal model in mind, but who knows only that cancer is the disease that her uncle had. We need to be somewhat careful in setting such cases up. To avoid begging the question about whether the concept is already being attributed in the very setup of the counterfactual situation, we should more neutrally describe her as, for example, sincerely asserting the sentence 'Cancer is the disease my uncle had', or evincing some other sort of behavior that makes the attribution plausible. The question now is whether this behavior is enough to license attributing to her the concept CANCER. We often attribute concepts to individuals even if they have rather exiguous capacities to identify instances or provide correct and comprehensive information about categories. It seems, on our normal ways of attributing thoughts and concepts, that Ada could have the concept CANCER but not have the causal model. Hence the concept CANCER cannot just be the causal model.

However, the concept cannot just be identified with the piece of information that Ada has either, since Beth only knows that cancer is something that smokers often get (or, to again avoid begging questions, she is only in a position to sincerely assert 'Cancer is something smokers often get'). And this is a different piece of information from what Ada possesses. So the concept cannot be identified with the information Ada represents, either. But it is clear that for exactly the same reason, neither can the concept be identified with the information Beth represents. Call this maneuver the *subtraction strategy*. It would seem that this strategy is one that can be iterated for any piece of information that a localist might think is constitutive of a concept. By appeal to the subtraction strategy, atomists claim that for any cognitive connection between concepts C and F it is possible that someone could have C but either (1) lack F entirely, or (2) not connect F cognitively with C. Since this can be iterated indefinitely, atomists conclude that the concept C

itself must be an unstructured representation to which all of these bodies of information can potentially become attached, but which isn't individuated by its links to them.

This strategy seems to be implicit in the atomist's insistence that concepts be portable across many different total doxastic states of a person. Margolis (1998b) uses the subtraction strategy to argue against Peacocke's (1998) view that conceptual content is partially constituted by a concept's relations to other contentful states that guide its application. Given the right circumstances, virtually any piece of information about a category, no matter how deeply held, could be rejected. We can imagine counterfactual cases in which we would say that people possess a concept (according to our normal attributive standards) but do not have the relevant relations to other concepts. Margolis calls this the 'Phenomenon of Abandoned Principles'. This echoes Quine's observation that potentially any link in our complex webs of belief can be severed. The subtraction strategy just involves looking at our hypothetical attributions of concepts to people in whom these links have been cut.

The subtraction strategy, however, falls short of establishing the strongest form of atomism. The strongest form of atomism is committed to the Strong Independence Principle:

(SIP): Possessing a concept does not require possessing any other concepts whatsoever.

These iterated subtractions, however, don't establish SIP. They show at best that no *particular* piece of information is necessary for concept possession. But each stage still requires that the concept-possessor have *some* other concepts. It doesn't follow from this that someone could have a concept despite having abandoned *all* other principles and the concepts used to formulate them. These other concepts are given in the examples as part of the subject's conception of the category picked out by the target concept. So the strategy itself seems to establish at best a Weak Independence Principle:

(WIP): Possessing a concept does not require possessing any other concepts in particular, although some concepts or other *are* required.

WIP is consistent with the subtraction strategy since at each stage we describe a new case of concept possession involving a different particular set of concepts. The strategy thus highlights our willingness to attribute concepts that are the same (in some sense) to individuals despite the diversity of ways they have of representing categories.

The unfolding of the strategy appears to depend on what I will call the Minimal Concept Possession principle:

(MCP): Someone has a concept C iff they have a representation of <C> that can play the appropriate causal/functional role in their cognitive economy.

According to MCP, concept possession requires at least the potential to represent a category in thought. The principle is underspecified, given that the particular causal/functional role that defines conceptual thought is challenging to spell out. Being able to combine with other representations to form complete propositional representations and being poised for use in reasoning and guiding behavior, including linguistic behavior, are undoubtedly a significant part of the appropriate role. Further conditions might be needed to distinguish conceptual ways of representing <C> from other ways the organism possesses, e.g., sensorimotor representations and the kinds of representation present at subpersonal or subdoxastic levels. The present formulation, vague though it is, will do in lieu of more explicit conditions distinguishing conceptual from nonconceptual representations.

Our justification for accepting the attributions used at each stage of the subtraction strategy derives from the Minimal Concept Possession principle: Ada, Beth, and the rest all

represent <cancer> in some appropriate way or other, hence each of them has a cancer-concept. The particular way they represent it, or how robust and detailed their conception of <cancer> is, does not seem to affect our willingness to attribute the concept to them. Atomists hold that our intuitions about concept attribution and possession support the claim that, for whatever constitutive links localists propose, a subtraction scenario can be created in which a subject is properly said to have the concept, but lack the links.

But this very fact about the subtraction strategy leaves the door open to the localist. Ada and Beth, she might say, simply have different cancer-concepts. Ada represents it under the cognitive content <disease my uncle had>, Beth does so under <something smokers get>, Cathy might do so by a complex content specifiable only by spelling out a causal model of the disease, Darla in yet another way, and so on. In light of the Minimal Concept Possession principle, each of them possesses a CANCER concept, since each has some representation with the R-content <cancer>. Given that the localist individuates concepts in terms of their referential and cognitive contents, these concepts differ along one semantic dimension. The appropriate moral to draw from the application of the subtraction strategy, then, is just that a theory of concepts should be *intersubjectively pluralist*: it should allow different subjects to possess concepts sharing their R-content but possibly differing in their C-content. Pluralism says that conceiving of an individual or category does not require representing it any particular way. Rather, there are many different ways in which individuals who can all conceive of the same thing might represent it. Atomists are right that MCP supports attribution of concepts with identical R-content across subjects with different conceptual resources and doxastic states. But from this true conclusion about the ability to refer, nothing follows about the structure of the vehicles that bear that reference. Given MCP, either all of these semantically structured vehicles are cancer-concepts, or none of them are.

Unless we assume that there is such a thing as *the* unique concept of cancer (or cats, or animals, or knowledge, or truth), we should not follow atomists in choosing the latter option.

A parallel case can equally be made for *intrasubjective* pluralism. First, an analogue of the subtraction strategy could be run using a single individual at different stages of her cognitive development: Ada at time t_1 versus Ada at time t_2, etc. But, second, the case for intrasubjective pluralism can be strengthened further by considering again the psychological models posited to explain how subjects represent and reason about categories. For no single cognitive model of concepts completely explains *all* of the relevant empirical data about people's performance. That this is true can be shown by a glance at the recent history of the psychology of concepts: insofar as there has been anything like progress in the discipline, it has involved shifts from the dominance of one kind of all-encompassing model to another, as new empirical results are found that the previously dominant models must struggle to explain. According to the standard story, then, the classical (definitional) theory gave way to prototypes and exemplars, which in turn gave way to the 'theory theory', and so on (see Komatsu, 1992 for review). No doubt this oversimplifies things radically, but in outline it is correct. None of the models that fall within each of these traditions is by itself a complete model of how our concepts function in even a few tasks or domains. But this is compatible with the claim that each model, where it is successful, has a true *partial* view of how the conceptual system works. If these models are correct depictions of how people sometimes represent the world in performing various tasks, then we are provisionally committed to the various representations that each of these models contains. No single unifying model seems poised to take the empirical high ground over its competitors. In order to capture the full range of human performance over a variety of contexts, tasks, materials, and domains, we might need to suppose that individuals typically have *many* readily available

concepts of a single category, corresponding to the particular kinds of concepts posited by these various different models.

On a fully pluralist view of concepts, in some tasks and situations subjects may represent cats by means of a set of exemplars (representations of particular experienced cats), on others they may do so by means of a causal model, on others they may employ prototypes, and so on. Counterfactual reasoning about a category, for instance, often depends on representing what the subject takes to be essential features of a category, or at least highly causally central ones, while discarding merely prototypical features (Keil, 1989; Rips, 1989). People can construct a bird-concept the cognitive content of which differs radically from their default bird-concept. So people, like the subjects in Rips' experiments, can accept that there might be a kind of bird that has the appearance, diet, and reproductive methods of a large insect, as long as such an unusual bird has important properties in common with birds (it has bird parents, or parts of its life-stage involve birdlike behavior). But these central properties may have nothing to do with how birds are conceived of in more ordinary situations. Prototypes or exemplars of birds may dominate in other reasoning situations. For instance, ordinarily we may be happy to judge that something is a bird if it is sufficiently similar to birds we can recall. Even single individuals may not conceive of birds in precisely the same way across two different situations.

The localism defended here, then, is committed to the following:

(Localism) Concepts have constituent structure.

(Dual Content) Concepts have both referential and cognitive content.

(Indiv*) Concepts are individuated by their referential and cognitive content.[13]

[13] The asterisk on 'Indiv*' is to mark its difference from the atomist principle Indiv.

In addition to these localist principles, we have seen that there is good reason to accept the pluralist thesis:

> (Pluralism) People typically have multiple concepts available for representing each possible referential content that they can conceive of.[14]

Pluralism holds both across and within subjects. Adopting pluralism entails accepting that that there is no such unique thing as *the* concept of an F. Given the Minimal Concept Possession principle, any appropriately situated representation of <C> can count as a C-concept. We can call the set of a subject's C-concepts her *family* of C-concepts. So pluralism amounts to the claim that typically, concept possession involves a subject's having a family of concepts rather than a single all-purpose representation deployed in all contexts.

Pluralism has several notable virtues. It allows us to endorse the intuitions elicited by the subtraction strategy without embracing atomism, since pluralistic localism is compatible with Weak Independence. It also accords with a substantial body of empirical results concerning the variety in our conceptual representations and their dependence on various aspects of task and context. Neither is pluralism a threat to the public shareability of concepts, for two reasons. First, many concepts are probably *de facto* widely shared and function as relatively stable default ways of relating cognitively to a domain. Support for this claim comes from the fact that developing successful cognitive models is an empirical enterprise requiring generalization across the behavior of many subjects. If a model is empirically well-validated, it is reasonable to think that this is so because the structures it posits are in fact had by many separate individuals. Second, even where concepts are not *de facto* shared, they are individuated only in terms of relatively

[14] Something is a *possible* referential content for a person in the event that he or she possesses a concept that has it as its reference.

small representational bundles. This renders them at least potentially shareable, unlike the holist's concepts, which are massively unlikely to be had by more than one person or person-stage.

6. Content and individuation

I now want to consider an objection to Indiv*, the principle that concepts should be individuated by their dual contents. Fodor (1998) argues that senses (which for present purposes are close enough to cognitive contents) do not carve concepts finely enough:

> Suppose I tell you that Jackson was a painter and Pollock was a painter, and I tell you nothing else about Jackson or Pollock. Suppose, also, that you believe what I tell you. It looks like that fixes the senses of the names 'Jackson' and 'Pollock' if anything could; and it looks like it fixes them as both having the *same* sense: viz. *a painter*. (*Mutatis mutandis*, it looks as though I have fixed the same inferential role for both.) Yet, in the circumstances imagined, it's perfectly OK—perfectly conceptually coherent—for you to wonder whether Jackson and Pollock were the *same* painter. […] So, then, by Frege's own test, JACKSON and POLLOCK count as different MOPs [modes of presentation]. But if concepts with the same sense can be different MOPs, then, patently, MOPs can't be senses. (p. 16)

Frege's test for sameness of MOP says, in one formulation, that *a* and *b* have different senses iff the claim that *a=b* would be informative for the subject. In Fodor's scenario, my uncertainty as to whether Jackson=Pollock means that I might simultaneously judge that Jackson is a drunk while judging that Pollock is not a drunk despite the fact that JACKSON and POLLOCK have the

same sense (A PAINTER). And this purportedly shows that concepts need to be individuated in terms of formal/syntactic features—they need to be carved as finely as forms of words.[15]

Putting 'cognitive contents' in for 'senses', a similar argument could be run against the position articulated here. But I don't think that this argument demonstrates the *necessity* for individuating concepts in terms of syntactic features. One possible strategy for the localist is to 'go linguistic' and allow that concepts can have forms of words as part of their cognitive contents. So, e.g., the concept one forms when one hears the word 'Jackson' might be: [PAINTER, CALLED 'JACKSON']. Since Pollock is introduced using a different name, the cognitive content of POLLOCK differs. So the concepts don't really match in their cognitive content after all. This amounts to agreeing with the basic claim that concepts should be individuated by forms of words (in public language, not Mentalese). One objection to this strategy is that it prevents cross-linguistic sharing of concepts: if my CAT concept contains CALLED 'CAT' and Pierre's CAT concept contains CALLED 'CHAT', then we do not share precisely the same cat-thoughts. While we might insist on strict shareability for some purposes, though, we needn't do so in general. Moreover, pluralism allows that concepts can differ in their cognitive structure within and across individuals. So the shareability objection has little force.

A more serious objection is that we need some criterion for individuating forms of words in public language. Two possibilities suggest themselves: (1) purely formal (phonological and orthographic) or (2) formal plus semantic. But adopting (1) will not invariably work. Suppose that on one occasion you hear me talk about a painter named 'Jackson', and you hear me do so

[15] Schiffer (1990) gives a similar argument against what he calls 'stereotypes' being modes-of-presentation: although I might have the same stereotype in mind for many distinct kinds, I can still entertain the possibility that those kinds are distinct. If modes of presentation are to account for cognitive differences, stereotypes cannot be modes of presentation. Of course, Schiffer ultimately doesn't agree with Fodor that syntactic properties can serve as modes of presentation either.

again on a later occasion. It seems that you can still coherently wonder whether Jackson=Jackson; that is, whether the first Jackson is the second Jackson. Here the cognitive content of each JACKSON concept is the same, viz. <painter, called 'Jackson'>. Something more than linguistically augmented cognitive content must account for the difference in cognitive value of these two concepts. One might attempt to carve forms of words more finely, so that 'Jackson' as heard on the first occasion is typed differently from 'Jackson' as heard on the second occasion. But this typing scheme lacks plausibility if we allow only phonological/orthographic features to individuate forms of words, since there seem to be no such differences here. Moreover, even if it is plausible for names, it is highly implausible for general terms like 'cat', and we need a general solution.

Adopting possibility (2) is equally problematic for typing public language terms as they are mentioned in cognitive contents. For in the present examples, the only meaning that a subject can plausibly attach to the introduced names is <a painter>. But then appealing to meanings adds nothing, since the introduced meanings are the same for each term. It seems that the linguistic strategy does not allow us to carve concepts finely enough in all cases. This isn't to say that cognitive content can't include linguistic information, though; only that there are cases in which concepts appear to be distinct despite having identical cognitive and referential content.

The moral is that content alone cannot individuate concepts in all cases. Other *non-formal* psychological properties might, though. To spell out this possibility I wish to borrow the notion of a *mental file*, which has been put to various uses by authors such as Perry (1980), Recanati (1993), and Millikan (2000). A mental file can be thought of as a discrete space in long-term memory, something like a reserved set of addresses in a computer's memory. Mental files function to store information that is (or is believed to be) about an individual, kind, or property.

They give structure to long-term memory, a structure that reflects possible topics of thought. So I might have a mental file in which my elephant-related information is stored, containing such facts as that elephants are grey in color, have wrinkled skin, have tusks and trunks, make a trumpeting sound, and are found in Africa and Asia, as well as in zoos worldwide. Insofar as the information in this file is all being predicated of the same thing, namely elephants, this is an elephant file. Mental files can be thought of as encyclopedia entries reflecting some portion of one's conception of their topic.

Ideally, information about one kind of thing doesn't spill over into other files. Filing systems shouldn't be needlessly redundant, particularly when the redundancy goes unnoticed by the file-keeper. But mistakes can happen. It might be that although I have an elephant file already, I later hear about a creature called a 'heffalump'. Not being sure what sort of things heffalumps are, I open a new mental file—a new address space in long-term memory—to collect information about them. At this point (since elephants are heffalumps) I actually have two elephant-files. There are two files here, not one, since files are distinguished by their distinct addresses, not just by their content. (Compare: we have the same manila file folder even if we re-label it and change its contents.) If I later discover this fact, I might merge the information into a single file, eliminating the potentially confusing redundancy.

While the notion of a mental filing system is rather abstract from the point of view of scientific theories of the organization of memory, it seems unproblematic to suppose that some such system exists, however it is implemented. And the existence of such a system of mental files can assist in overcoming Fodor's objection to individuating concepts solely by cognitive content and referential content. The elephant/heffalump case just canvassed provides the template for our solution. Suppose that I hear Fodor's story about Jackson and Pollock. I may

then form two concepts (two tokens) with the same simple cognitive content, <a painter>. But given that I have heard two different names used (or perhaps have heard the same sounding name on different occasions), I may open two separate mental files to keep track of information about each individual, supposing that they are or might be distinct. If that assumption later proves wrong, I am always free to merge the files. The fact that there are *two* files reflects the fact that I am uncertain whether the two cognitively type-identical concepts I have formed are genuinely co-referential. Where we are uncertain whether two concept tokens co-refer, we run the risk of failing to view tokens of the same concept type *as* such. The fact that each of these concept tokens is attached to a separate mental file lends them distinct functional roles. And this difference in functional roles is reflected in the fact that I might think different predicative thoughts using one of these tokens than using another. Being attached to a particular mental file provides a non-formal psychological property of concepts that does the job needed to answer Fodor.

A larger point can be drawn out from this picture. Notice that a person can possess two distinct tokens that have identical C-content and R-content but which differ in their functional roles due to their being attached to distinct mental files. It follows that content does not exhaustively determine functional role for concept tokens. So Fodor's parenthetical comment about having fixed identical functional roles for the newly introduced concepts is incorrect. This goes beyond the point made above about featural structure (hence cognitive content) being insufficient by itself to determine functional role. There, the issue was that we need to fix a set of psychological processes in order to fix a concept's role. Here, the point is that we may also need to consider its relations to other resources such as mental files. Content and functional role are at least partially separate parameters of a representation.

7. Analyticity deflated

A final objection to positing semantically structured concepts is that doing so commits us to an analytic/synthetic (a/s) distinction, since it is committed to conceptual connections arising from the part-whole structure of concepts. If concepts are frames, then many concepts literally contain other concepts. So, necessarily, having some concepts entails having certain others. Both Kant and Frege have articulated a view of analyticity as conceptual containment (although both also adopted other views of analyticity as well). On this view, the thought that Fs are Gs is analytic iff the concept G is a constituent of the concept F. If Waldo conceives of cats under the concept [FURRY, LIKES MILK, GOOD AS PETS], then it is analytic for Waldo that cats like milk. So localism commits itself to analyticities, on one widely held understanding of analyticity. A pluralist form of localism seems only to multiply this problem, since for every F-concept a person possesses, a new set of analytic entailments concerning Fs is generated. The greater the degree of pluralism, the greater the burden of commitment to analyticities.

Moreover, the problems of the present view would seem to be even *worse* than those usually associated with analyticity. Traditionally, analytic statements were held to express necessary truths. But on the present view, cognitive content and referential content can (and often do) diverge from one another. Cognitive content might not even be *true* of referential content, in the case of inadequate concepts, let alone necessarily true. So it seems there can potentially be *false* 'analytic' judgments. And as the scare quotes indicate, this conclusion seems intolerable.

Traditionally, the notion of analyticity was intended to explain a number of disparate metaphysical and epistemological phenomena. Analytic sentences can be known to be true just in virtue of knowing what they mean. Since knowledge of meaning was supposed to be *a priori*,

they were knowable independent of knowledge of the world; hence, analytic sentences were supposed to be knowable *a priori*. They were also supposed to be necessary—indeed, analyticity was intended to *explain* necessary truth. Commitment to the full burden of the traditional a/s distinction, particularly the linguistic theory of necessary truth, would be onerous. But commitment to some form of analyticity or other need not be. Localists can readily accept a defanged, psychologized notion of analyticity. (The most careful discussion of how a psychologized notion of analyticity might work has been developed by Lyons (2005); at several points my discussion here will follow his.)

First, commitment to conceptual containment relations is motivated not by the metaphysical and epistemological theses characteristic of the traditional notion of analyticity, but rather by the explanatory demands of cognitive psychology. In virtue of being committed to the ontology expressed in the best cognitive models of concepts, what the localist is committed to is that concepts typically contain other concepts. In virtue of the existence of conceptual structure, some judgments become accessible to a thinker just in virtue of her ability to 'inspect' the constituents of her concepts. If a complex concept C contains feature F, then she is in a position to judge that Cs are F in virtue of taking note of this containment relation. All that is needed is the following set of cognitive processes:

> (1) an operation that *detaches* features from frames, essentially the opposite of the operation that combines concepts into new frames;
>
> (2) the ability to form propositional judgments by linking concepts in subject and predicate roles;
>
> (3) the rule that if F can be detached from C, then it is permissible to form the judgment that C is F.

Processes (1)-(3), which are sensitive only to containment relations among concepts, are sufficient for entertaining new judgments derived solely from conceptual structure. The resulting judgment is one in which, to borrow the classical formulation, the predicate concept is contained in the subject concept. These judgments are 'representationally analytic' in the sense laid out by Lyons.

Analytic judgments formed by applying processes (1)-(3) do not involve accessing any other beliefs that the subject might have. In particular, they don't involve accessing any separately stored empirical beliefs. In a certain sense, this means that these judgments are ones that can potentially be formed without appealing to any knowledge of the world. This makes them a priori, in at least one sense of the term: their justification comes from the rule expressed in (3) above, not from any empirical information. (Of course, this rule need not be explicitly represented in order for such judgments to be formed.) Even if these analytic judgments are a priori in this sense, though, the subject might not *know* that they have this status. For processes (1)-(3) don't necessarily involve consciousness or awareness at all. One may use concepts without being conscious of their internal structure. And one may form judgments that are derived from conceptual structure without necessarily being aware that this is what one is doing. There need not be any obvious mark of the inferential origin of a particular judgment that would indicate its being derived from conceptual structure, since access to the precise processes that produce our thoughts can be indirect even for the one who possesses those thoughts.

What of the connection between analyticity and necessary truth? Suppose that Paul has a PENCIL concept that includes the features [USED FOR WRITING, CYLINDRICAL, CONTAINS LEAD]. Then given only the operations mentioned above, he can form the judgment that pencils contain

lead. Since concepts have two contents, so too do the thoughts that they compose. The contents of this judgment can be represented as:[16]

(1_C) <<<used for writing> & <cylindrical> & <contains lead>> → <contains lead>>

(1_R) <<pencil> → <contains lead>>

Here, (1_C) gives the C-content of the judgment and (1_R) gives its R-content. Proposition (1_C) is clearly a logically necessary truth: it is expressed in virtue of the formal structure of the judgment itself. Proposition (1_R), however, is not a necessary truth. We can see this by considering worlds in which pencils do not contain lead. Perhaps pencils could be devised that don't make marks by scraping a lead across paper, but instead spray a focused graphite mist onto the page. These devices might be correctly thought of as leadless pencils. If so, then (1_R), while actually true, could be false. So the judgment that pencils contain lead expresses two propositions, one of which is a necessary truth underwritten by logical form, the other of which is a contingent truth. (1_C) might be thought of as a conceptually necessary truth. But insofar as truths like (1_C) are derived only from the logical or structural properties of concepts, they are innocuous. Even on atomistic views, concepts can combine into propositional structures, and these structures will underwrite exactly similar structural 'analyticities'; e.g., <<<brown> & <cow>> → <cow>> (see Fodor & Lepore, 2002).

The same two-level semantic apparatus helps to address the worry about false 'analytic judgments'. Suppose Carol has an inadequate concept of coral. According to her, coral is a strange kind of underwater plant. If she entertains this judgment, again she expresses two propositions (I omit other features that might be part of her coral concept):

[16] I assume that the frame structure itself contributes the conjunctive content displayed here.

(2_C) <<kind of plant> → <kind of plant>>

(2_R) <<coral> → <kind of plant>>

Once more, (2_C) is like (1_C) in being a logically necessary truth. But (2_R) is not even contingently true, as (1_R) was; rather, it is false. At the level of R-content a judgment may express a falsehood despite expressing a necessary truth at the level of C-content. If we regard the level of cognitive content as the appropriate level to locate a psychologized notion of analyticity, then there cannot in fact be false analytic judgments. All analytic judgments express necessary C-propositions, although some may also express false R-propositions.

The traditional notion of analyticity had the effect of dividing what a subject knows about a category into two bodies of information: the information that is constitutive of her possessing the concept of that category, and the merely collateral information she possesses about it (often called her conception of it; see Woodfield, 1991). On pluralism, though, the concept/conception distinction becomes simultaneously more complex and, peculiarly, less important. With respect to each of the concepts in a family, we can distinguish between what is constitutive of having *that* F-concept and what is collateral information relative to it.

But it is perfectly possible—indeed, likely—that information which is constitutive of one F-concept (e.g., a category exemplar) will be collateral relative to another F-concept (e.g., a causal model), and vice-versa. So there is no *absolute* fact of the matter about whether some information is constitutive of possessing a concept generally; since concept possession usually involves families of F-concepts, it all depends on the particular F-concept that we are focusing on. A piece of information may partially constitute one of a person's F-concepts but not another. Moreover, the distinction between concept-constituting information and so-called background knowledge, or a person's conception of a category, also becomes relativized. Whether

information is part of the background or not depends on the specific F-concept that we are considering. Again, relative to one such concept a body of information might be merely a set of background facts, while relative to another it might be concept-constituting. There is no absolute sense in which any information is constitutive of a concept or merely part of one's conception. So a localist need not be committed to placing any serious weight on the different psychological roles of concept-constitutive vs. collateral information.

8. Conclusions

The argument given here for semantic localism turns on two things: the prevalence of structured representations in empirical theories of concepts, and the plausibility of giving those structures a certain semantic interpretation. I think that we should take the best models of concepts and central conceptual phenomena at face value, as our best guide to what our concepts are and how they work. But the cognitive models themselves do not *determine* that the semantic interpretation I have given here is correct. Psychologists do not often concern themselves with talk about truth conditions, and certainly other possible interpretations of the same models and data could be put forward. What I have tried to establish is that this interpretation is coherent, defensible, and useful in overcoming standard atomistic objections. If we accept the reality of the representations modeled by psychologists, we can give them a semantics that overcomes many of the standard objections to localism. As I noted at the outset, atomism is frequently defended as a position of last resort: we must accept it, because no alternatives succeed. This is not a bad argumentative strategy, so long as all of the available alternatives have been examined. I've tried to show that there is somewhat more space available on the side of non-atomistic theories. Pluralistic localism can succeed where other localistic theories can't. This renders it worthy of further attention and consideration.

References

Anderson, J. R. (1978). Arguments concerning representation for mental imagery. *Psychological Review, 85*, 249-277.

Aydede, M. (1998). Fodor on concepts and Frege puzzles. *Pacific Philosophical Quarterly, 79*, 289-294.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211-227.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization* (pp. 21-74). Hillsdale, NJ: Lawrence Erlbaum.

Block, N. (1993). Holism, hyper-compositionality, and hyper-analyticity. *Mind and Language, 8*, 1-26.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and Conceptual Development* (pp. 141-174). Cambridge: Cambridge University Press.

Burge, T. (1979). Individualism and the mental. In P. French, T. Euhling & H. Wettstein (Eds.), *Midwest Studies in Philosophy* (Vol. 4, pp. 73-121). Minneapolis: University of Minnesota Press.

Burnett, R. C., Medin, D. L., Ross, N. O., & Blok, S. V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology, 59*, 3-10.

Chalmers, D. J. (2002a). The components of content. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 608-633). Oxford: Oxford University Press.

Chalmers, D. J. (2002b). On sense and intension. *Philosophical Perspectives, 16*, 135-182.

Chalmers, D. J. (2003). The nature of narrow content. *Philosophical Issues, 13*, 46-66.

Fillmore, C. J. (1968). The case for case. In E. Bach & Harms (Eds.), *Universals in Linguistic Theory* (pp. 1-88). New York: Holt, Rinehart, and Winston.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (1989). Substitution arguments and the individuation of belief. In G. Boolos (Ed.), *Method, Reason, and Language*. Cambridge: Cambridge University Press.

Fodor, J. (1990). A theory of content II. In *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fodor, J. (1994). Concepts: A potboiler. *Cognition, 50*, 95-113.

Fodor, J. (1998). *Concepts*. Oxford, UK: Oxford University Press.

Fodor, J., & Lepore, E. (1996). The pet fish and the red herring: Why concepts still can't be prototypes. *Cognition, 58*, 243-276.

Fodor, J., & Lepore, E. (2002). Why meaning (probably) isn't conceptual role. In *The Compositionality Papers* (pp. 9-26). Oxford, UK: Oxford University Press.

Gelman, S. A. (2003). *The Essential Child: The Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition, 38*, 213-244.

Higginbotham, J. (1998). Conceptual competence. In E. Villanueva (Ed.), *Concepts* (pp. 149-162). Atascadero, CA: Ridgeview.

Keil, F. C. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112*, 500-526.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lepore, E., & Loewer, B. (1986). Solipsistic semantics. In P. French, T. E. Uehling, Jr., & H. K. Wettstein (Eds.), *Midwest Studies in Philosophy: Studies in the Philosophy of Mind* (Vol. 10, pp. 595-614). Minneapolis: University of Minnesota Press.

Lyons, J. C. (2005). Representational analyticity. *Mind and Language, 20*, 392-422.

Margolis, E. (1998a). How to acquire a concept. *Mind and Language, 13*, 347-369.

Margolis, E. (1998b). Implicit conceptions and the phenomenon of abandoned principles. In E. Villanueva (Ed.), *Concepts* (Vol. 9, pp. 105-114). Atascadero, CA: Ridgeview Publishing Co.

Markman, A. B. (1999). *Knowledge Representation*. Mahweh, NJ: Lawrence Erlbaum Associates.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge: Cambridge University Press.

Millikan, R. G. (1998). A common structure for concepts of individuals, stuffs, and basic kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences, 22*, 55-65.

Millikan, R. G. (2000). *On Clear and Confused Ideas*. Cambridge, MA: Cambridge University Press.

Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Peacocke, C. (1998). Implicit conceptions, concepts, and rationality. In E. Villanueva (Ed.), *Concepts* (Vol. 9, pp. 43-88). Atascadero, CA: Ridgeview Publishing Co.

Pereboom, D. (1995). Conceptual structure and the individuation of content. *Philosophical Perspectives, 9*, 401-426.

Perry, J. (1980). A problem about continued belief. *Pacific Philosophical Quarterly, 61*, 317-332.

Prinz, J. (2000). The duality of content. *Philosophical Studies, 100*, 1-34.

Prinz, J. (2002). *Furnishing the Mind*. Cambridge, MA: MIT Press.

Putnam, H. (1975). The meaning of 'meaning'. In *Mind, Language, and Reality: Philosophical Papers* (Vol. 2, pp. 215-271). Cambridge: Cambridge University Press.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Malden, MA: Blackwell.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27*(5), 709-748.

Rehder, B. (2003b). A Causal-Model Theory of Conceptual Representation and Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1141-1159.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 21-59). New York: Cambridge University Press.

Robbins, P. (2002). How to blunt the sword of compositionality. *Nous, 36*, 313-334.

Schiffer, S. (1990). The mode-of-presentation problem. In C. A. Anderson & J. Owens (Eds.), *Propositional Attitudes: The Role of Content in Logic, Language, and Mind* (pp. 249-268). Stanford, CA: CSLI Publications.

Simons, D. J., & Keil, F. C. (1995). An abstract-to-concrete shift in the development of biological thought. *Cognition, 56*, 129-163.

Sloman, S. (2005). *Causal Models*. Oxford, UK: Oxford University Press.

Sloman, S., Love, B. C., & Ahn, W.-k. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.

Taylor, K. A. (1993). Narrow content functionalism and the mind-body problem. *Nous, 23*, 355-372.

White, S. (1982). Partial character and the language of thought. *Pacific Philosophical Quarterly, 63*, 347-365.

Woodfield, A. (1991). Conceptions. *Mind, 100*, 547-572.