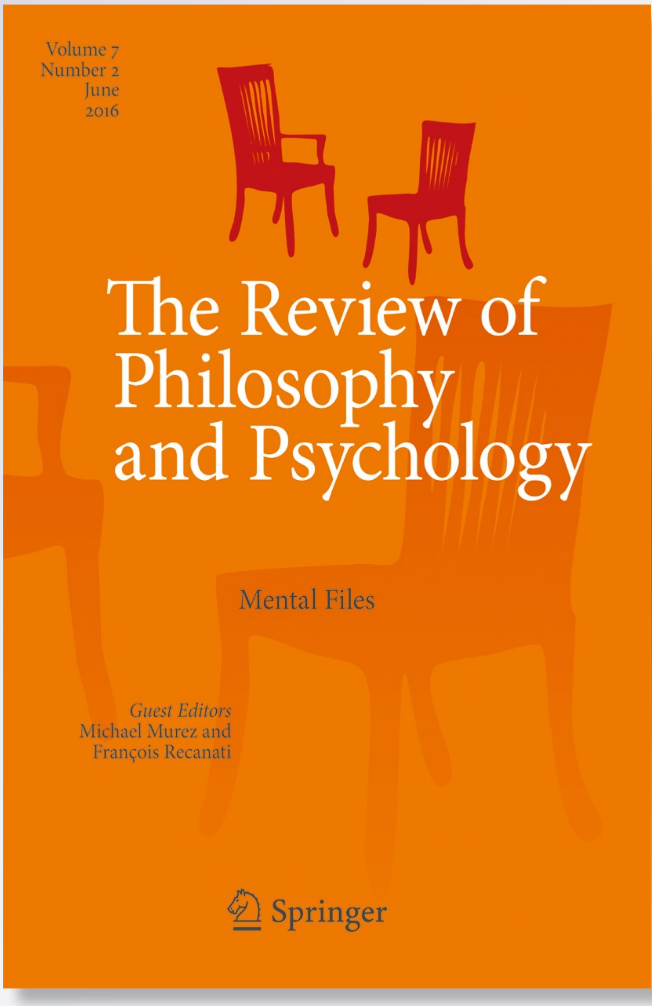
Mental Graphs

James Pryor

Review of Philosophy and Psychology

ISSN 1878-5158
Volume 7
Number 2

Rev.Phil.Psych. (2016) 7:309-341
DOI 10.1007/s13164-015-0280-1



Rev.Phil.Psych. (2016) 7:309–341
DOI 10.1007/s13164-015-0280-1

Mental Graphs

James Pryor[1]

Published online: 24 September 2015

Abstract I argue that Frege Problems in thought are best modeled using graph-theoretic machinery; and that these problems can arise even when subjects associate all the same qualitative properties to the object they're thinking of twice. I compare the proposed treatment to similar ideas by Heck, Ninan, Recanati, Kamp and Asher, Fodor, and others.

1

Models of the attitudes span a range from the not-so-structured — merely sets of hypotheses compatible with what the agent Vs — to commitment to particular mental symbols, and access to as much structure as they may have.[1] This paper identifies interesting intermediate models, that posit *some* structure but are noncommittal about how it's concretely realized in our cognitive architecture. For some explanatory purposes this seems a preferable level of abstraction. States as described by these models can then be shared between creatures who realize that structure differently.

The kind of structure I target is not Boolean structure, such as a difference between believing `P` and believing `not-P` $\supset$ `P`. Rather I aim to model the property attitudes have when they purport to be about a common object (or common sequence

[1]Examples of the latter extreme are: Fodor 1975, 1978, 1990; Field 1978; Schiffer 1981; Moore and Hendrix 1982; Richard 1990; Crimmins and Perry 1989; and Crimmins 1992.

✉ James Pryor
jim.pryor@nyu.edu

1 Department of Philosophy, New York University, New York, NY, USA

of objects), as when Hob believes that someone$_i$ blighted Bob's mare, and wonders whether she$_i$ killed Cob's sow.[2]

The framework I'll present has some resemblance to what Recanati and others have formulated in terms of "mental files," the topic of this volume. What follows could be understood as one way to (abstractly) develop some of their framework. I'll address some possible differences in Section 8. My strategy even more closely resembles Ninan's "multi-centered" view of mental content[3]; and as I'll explain, there is an underlying unity with Heck's 2012.

A deeply-entrenched idea in theorizing about mental states is that they can be individuated by the pair of their *content* plus the *attitude* we hold towards it. What exactly that idea requires is unclear, because (at least) what counts as *content* is up for negotiation. But despite some pockets of readiness when externalism is discussed to allow that identity of content may be non-transparent, I think most philosophers are reluctant to allow thoughts with the same content, or the contents themselves, to differ merely numerically, as Max Black's spheres are alleged to differ. A subsidiary aim of this paper is to join Heck 2012 in encouraging more sympathy for that possibility, one that my framework is well-suited to model.[4] On some natural ways to talk of "content," we could call this the possibility of believing the content that P (simultaneously) twice, rather than once. I'll present cases I find to speak most compellingly for this in Section 8. The earlier parts of this paper are devoted to developing the models that can even represent the idea.[5]

[2] As this familiar example hints, the phenomena we'll be considering don't always require the existence of an external referent, and can plausibly exist between the attitudes of different agents. But we'll postpone thinking about reference-failure until later, and won't address the interpersonal cases here at all.

[3] See Ninan 2012 and 2013. Versions of this view have also been proposed or engaged with by Austin 1990, pp. 43–9 and 136–7; Spohn 1997; and Chalmers in several places, such as his 2003, p. 228; 2011, pp. 625–7; and 2012, pp. 286–7, 397 note 7. Ninan develops the idea furthest. In the versions discussed by Austin and Chalmers, the additional centers are occupied by experiences or regions of one's sensory field, not by external objects. (Austin doesn't have sympathy for these views.) Torre 2010 and Stalnaker 2008 (pp. 73–74) propose a multi-centered view of communication but not of individual thought. See also Ninan's comparisons of his view to Hazen 1979 and to Lewis 1983a (esp. p. 398) and 1986 Chapter. 4.4.

[4] Ninan's and Recanati's theories might also represent this, though they don't discuss the issue directly. I'll discuss Recanati's stance towards this in Section 8.

When I say "differ merely numerically," I permit differing also in terms of relations that must themselves be specified partly numerically; see example (34) below.

[5] Views that commit to particular mental symbols can also make sense of this possibility, by just positing that a mental sentence appears in an "attitude box" multiple times. (Fodor for example proposes to ground cognitive differences in numerical differences between mental particulars: see his 1994, pp. 105–9 and 1998, pp. 16–17, 37–39. See also Richard 1990, pp. 183–5; and the treatment of "beliefs as particulars" in Crimmins and Perry 1989, pp. 688–89, and Crimmins 1992.) But like Heck (see esp. his notes 34 and 35), I'm aiming for a more abstract account.

Views that think of attitudes as fundamentally ternary, involving relations to both a "content" and a "guise," might also *seem* to be amenable to this possibility. But like Heck (see especially his pp. 144–49), I'm inclined to use "content" to include all the non-haecceitistic nature of a state that plays a role in intentional explanation. This approximates Loar 1988's use of "psychological content" — though Loar's substantive internalist *theory* of psychological content should not be presupposed. Understanding "content" this way, the ternary theorist's "guises" *should* count as part of a state's "content," and so these theorists aren't yet envisaging a possibility as radical as Heck and I are proposing. See also the discussion of neo-Fregeanism in Sections 8 and 9 below.

2

Alice believes her local baseball team has several Bobs on it. In fact, they only have one Bob. Maybe this happened because Alice met that same player repeatedly, and didn't realize she was getting reacquainted with someone already familiar. Or maybe she only met him once, but through some cognitive glitch she mistakenly opened several "mental files" for that single player. Either way, let's suppose she has now forgotten the specifics of how she met "each" Bob; and let's suppose that the memories, information, and attitudes she associates with "each" of them is the same — or as close to this as possible. (We will discuss how close it is possible to get later.)

Presently, Alice is thinking how the voting for her team's Most Valuable Player will go. There's a thought she can entertain that we might articulate, on her behalf, like this:

(1) Bob_i will vote for Bob_j.

where $i \neq j$.[6] The indices are meant only to be suggestive; we'll have to work out what it amounts to for Alice to have such a thought.

Even with only this suggestive start, already we can formulate an interesting question, whether Alice would also be in a position to entertain a distinct thought, articulable like this:

(2) Bob_j will vote for Bob_i.

The story I've told certainly doesn't put Alice in a position *to have more justification for* accepting one of these thoughts over the other; but that doesn't settle whether there are two distinct thoughts available here for her to entertain. (If we think there aren't, perhaps the way she'd have to express her single thought is One of the Bobs will vote for the other.)

Whatever our views on that question, though, most will want to distinguish the thought we've articulated with (1) from a thought articulable like this:

(3) Bob_i will vote for Bob_i.

It's only this last thought whose acceptance we'd expect to rationally commit Alice to the thought that anyone self-voted. And in my story, Alice doesn't seem to be so committed. Yet she was only ever thinking of a single player throughout. It's just that when thinking (1) — but not when thinking (3) — she'd unknowingly be thinking of this single player *twice*, or via two different presentations.

This is a familiar contrast; but there is some awkwardness finding a vocabulary perfectly suited to characterize it.

One phrase I just employed was "unknowingly thinking of a single player twice." But how should we describe the contrasting case (3)? Shall we say that there Alice *knowingly* thinks of a single player twice? That over-intellectualizes what's going on in case (3). Nothing I said required Alice to have any opinion about her own thoughts; and absent some controversial assumptions about the transparency or luminosity of

[6]She assumes that players on the team are allowed to vote; they are even allowed to vote for themselves. More precisely, I'll treat will vote for as an extensional binary predicate that's neither reflexive nor irreflexive.

introspection, it's not obvious that if she *did* have opinions about her own thoughts, they'd have to be correct.

Another phrase I employed was "two different presentations." That might be okay if we understand it to mean two *numerically* different presentations; but it's too easily heard as implying that the presentations must differ in some *qualitative*, that is, not-merely numerical, way. And that further step should be a substantive one. As I mentioned above, I will be arguing that it should sometimes be resisted: Alice's two presentations in case (1) needn't differ *except* numerically. But even if I'm wrong, the fact that this can be intelligibly disputed speaks for favoring a vocabulary that doesn't prejudice the question.

One recently popular way to characterize the difference between cases (1) and (3) is to say that in the latter but not the former, Alice thinks about Bob in a way that *coordinates* the two argument roles her thought ascribes to him.[7] This can also be misconstrued, but it's the vocabulary I will adopt.

Fans of structure are willing to count thoughts as distinct even when their contents are "logically equivalent." One form this can take is to say, here are different contents for thoughts to have:[8]

(1′) $(\lambda x.\, x$ will vote for Bob) Bob

(3′) $(\lambda x.\, x$ will vote for $x)$ Bob

If one's sympathetic to that, then one might describe our case (1) as one where even if Alice thinks contents of form (1′), and others like it, she doesn't think contents of form (3′). In case (3), she *does* think contents of form (3′).[9] That characterization

[7] This phrase is borrowed from Fine 2007. One wrinkle is that Fine's first uses of this phrase are to discuss theories where two *variable* occurrences have a *range of values* as their interpretation, and the semantics requires *those ranges* to be "coordinated": that is, choosing a value from the range for one variable occurrence fixes what value the other takes. As Fine acknowledges, that picture doesn't carry over to the semantics of names, which don't have such ranges of values. (See his pp. 23–4, 29–31, 39–40.) But Fine does anyway talk of names and thought components also being "coordinated." His paradigms for this are sentences or thoughts like the ones in our case (3), as opposed to our case (1).

In borrowing Fine's vocabulary, I don't commit here to any of the specific semantic proposals he defends. Moreover, the present paper solely concerns models of thought, not language. I'll discuss Fine's proposals about language in detail elsewhere.

[8] In addition to works cited in note 9, below, see also Wiggins 1976a, pp. 164–6; Wiggins 1976b, pp. 230–1; Soames 1985, note 12; Salmon 1989, pp. 269ff; Salmon 2006c; Salmon 2010; and Higginbotham 1991, pp. 360–1.

[9] There is a cluster of views, in linguistics associated most strongly with Reinhart 1983, and in philosophy with Salmon 1986a; see also Salmon 1992; Kaplan 1986, pp. 269–71; Soames 1987a, pp. 220–23 and note 24; Soames 1987b, esp. Section 8; Soames 1989/90, pp. 204ff; McKay 1991; and Bach 1987, pp. 256–7. (McKay's is the most linguistically sophisticated of these philosophical treatments. The rest of Bach's Chapter. 11–12 argue against extending the strategy beyond reflexive pronouns.) The specific commitments of these authors differ, but they share a strategy of explaining the linguistic (and sometimes cognitive) phenomena associated with expressions of coordinated thought always in terms of variable binding, as in (3′). It's not essential to this view that one think that (3) *itself* has the logical form of (3′) — indeed most of these authors do not. We would however naturally expect a subject who'd express her belief with (3) to also have beliefs of form (3′). So this encourages the idea that variable binding is diagnostic of the phenomena we're studying, or at least is the dominant phenomenon in the neighborhood. Reinhart's views have been very influential in the study of anaphora; but see Heim 1998 for a survey of its problems. I will explore the connections between the linguistics and the philosophical side of this family of views in other work.

doesn't seem to commit you either way on the question whether Alice's two thought-tokenings of Bob in (1) *themselves* have different contents. And so it might seem an attractively minimal way to capture the difference illustrated by our cases (1) and (3). (If you wanted to go on to distinguish case (1) from *case (2)*, then you'd have to say things less minimal.)

I'll defer for another discussion a detailed consideration of how much explanatory coverage the idea just sketched can achieve. I'll just announce my conviction that the phenomena we've been intuitively characterizing extend beyond cases where they correlate with variable binding structure. To take just one example:

(4a) Carol will vote for Bob, but Bob won't.

is best understood as having the binding structure of:

(4b) $(\lambda x.\ x$ will vote for Bob) Carol, but $\neg$(_____ Bob)

where the _____ is occupied by the same predicate expressed by the lambda term on the left.[10] Rather than:

(4c) $(\lambda x.\ x$ will vote for Bob) Carol, but $\neg$ $((\lambda x.\ x$ will vote for $x)$ Bob)

or:

(4d) $(\lambda y.\ (\lambda x.\ x$ will vote for $y)$ Carol, but $\neg$(_____ y)) Bob[11]

Yet, the thought's having the structure of (4b) should intuitively still be compatible with its being a "coordinated" way of thinking about Bob — one where the thinker uses only a single mental presentation for him.

So in this discussion, I will assume that the difference between coordinated and uncoordinated[12] thinking can't (always) be attributed to the binding structure of the predicates one's thoughts ascribe. Instead, I will sketch a different way to represent this difference as grounded in the structure of one's thinking.

[10] See the discussion of examples 49, 53, and 55 in Evans 1980; examples 13 and 20 in Heim 1998; and Pinillos 2011, p. 312.

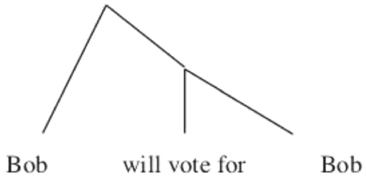
[11] See Soames 1994, top of p. 260. This is not a plausible syntactic structure for (4a).

[12] We should respect the difference between a theoretical model (i) representing a subject as *not* coordinating certain argument places, and (ii) merely being silent (in whole or in part) about what argument places are coordinated. I use the term "uncoordinated" in sense (i). Fine sometimes uses "uncoordinated" in sense (ii) (2007, pp. 52, 56–9, 77); but often uses it in sense (i) (pp. 55, 69, 78, 83, 96, 111, 117), and that is how I am using it here. This is what Fine also calls "negatively coordinated" (p. 56) and Salmon 2012, p. 437 note 40 calls "withholding coordination." Yet a further notion would be (iii) representing a subject as taking certain argument places to *definitely be occupied by different objects*. As far as I'm aware (and contra Salmon 2012, p. 409), no extant theory of coordinated thinking makes use of notion (iii).

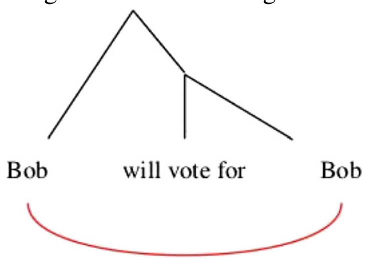
3

I've already declared that the models I'll be developing are abstract, that is, less committal about their cognitive implementation than some alternatives. What's also true is that I'll only be presenting a *general framework* for modeling a subject's thoughts. As we'll see, the framework can be specifically deployed in different ways. Some of these ways are more finely structured (more "linguisticy"); others less finely structured (more "logicy"). Which of those variants is more appropriate will depend on the explanatory work the model is supposed to perform.

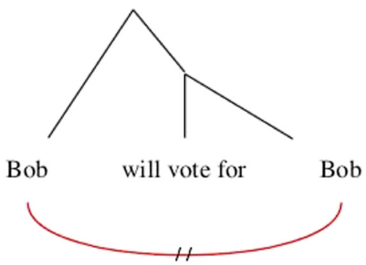
To get us started, though, it may help to sketch a fine-grained cousin of my view. This begins with the idea that thought contents have predicative and logical structure, as championed by King and others.[13] On their views, the content of Alice's thought in case (1) would be something like the syntactic structure of the sentence we use to articulate it, except with the leaves labeled by objects and relations, rather than by expression-types:



To that base, we add the further refinement of "wires" connecting the two occurrences of Bob when Alice is thinking a coordinated thought:



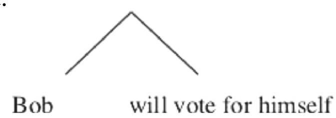
and the absence of such "wires" (or "unwires") when Alice is thinking an uncoordinated thought:



[13] I take King 1996 and 1995 as paradigmatic. This tradition goes back to Lewis 1970 and Cresswell and von Stechow 1982; and has roots in Carnap's notion of "intensional isomorphism." See also the later developments of King's view, and comparison to other accounts of structured propositions, in King 2007 and 2011.

(Alternatively, one could tag all the leaves with indices, understanding leaves to be wired when they have the same object and index, and unwired when they have the same object but different indices.)

A natural question that arises is what is the relation between the content in the wired tree and the content:



Are they the same, at least in this simple case? Does either entail the other? Those are good questions; but we will not try to answer them now.

If a view of this sort allowed "wires" and "unwires" among *all* of the occurrences of Bob within *and across* all of Alice's beliefs (and also her other attitudes), then it would turn out to coincide with the finest-grained versions of the general framework I'll be proposing.[14]

However, for many explanatory purposes we'll want a notion of "content" that abstracts from some syntactic detail. For example, we may want to count the thoughts articulated by:

(5a) Bob will vote for Carol; and Carol will vote for Bob.
(5b) Carol will vote for Bob; and Bob will vote for Carol.
(5c) Bob and Carol will vote for each other.

[14] Such "wires" have been deployed in different ways in philosophy and linguistics, which should not be confused. First, there was Quine's suggestion in his 1940 that they could be used in place of bound variables. This suggestion is repeated in Kaplan 1986, p. 244; Salmon 1986b, p. 156; and Soames 1989/90, p. 204. See also Evans 1977, pp. 88–96. (On p. 102, Evans permits the wires to cross sentence boundaries, and from pp. 104ff, they're also used to join "donkey pronouns" to their antecedents. Here the notation must be interpreted differently.) King 2007 applies Quine's idea to his structured propositions, which have a language-like structure; see his pp. 41–2 and 218–22. One advantage of this device is to abstract away from the alphabetic identity of variables. (Another technique for securing that same advantage, commonly used in writing compilers, can be read about at https://en.wikipedia.org/wiki/De_Bruijn_index.)

The second way in which "wires" have been used is to join positions in a proposition that are *already occupied* by objects, as a way to distinguish between recurrences of those objects that require or involve coordinated thinking and those that don't. Soames 1987b mentions this idea at p. 112, attributing it to unpublished work by Kaplan; and it is appealed to in Fine 2007, at pp. 54–7 and 77–8. See also Fine's notions of "token individuals" and "token propositions" in Fine 2010, pp. 479–80; and see Salmon 1986b, p. 164; Salmon 1992, p. 54–55; and Pinillos 2011, p. 319. Richard 1990 and "ILF" theorists like Larson and Ludlow 1993 posit other mechanisms that can also induce this structure. This is the use of "wires" I was invoking in the text.

The third way in which "wires" (or similarly capable mechanisms) are deployed is in some linguistic work on anaphora. For example, see Higginbotham 1983 and 1985; Moltmann 2006, pp. 236ff; Heim 1998; and Fiengo and May 1994. (See also Soames 1989/90, pp. 206ff and note 14; McKay 1991, pp. 724ff; and the usage in Evans 1977 from pp. 102ff, mentioned above.) These authors aren't working with a single notion, but there are notable similarities. Unlike the first use of "wires," the relations they posit are not restricted to positions that could be bound by a quantifier. ("Donkey pronouns" can also be joined to their antecedents.) Additionally, many of these authors insist that the relations they're positing aren't transitive or Euclidean; whereas in the first and second uses of "wires", we typically see equivalence relations, at least with the contents deployed by a single subject on a single occasion (Pinillos is an exception). I believe there are connections between the work done by these third authors and what the second use of "wires" is trying to implement; but those connections need to be developed and defended. We shouldn't *begin with* the assumption that the similar-looking diagrams are representing the same thing in each case.

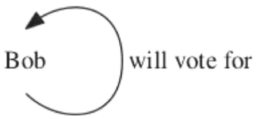
as having the same (coordinated) contents. We may want to count thoughts like these:

(6a) Bob was afraid to vote.
(6b) Bob feared voting.
(6c) Voting frightened Bob.

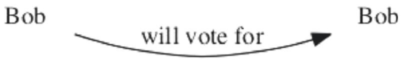
as having the same contents, too, despite their varying syntax. The general framework I'll propose can also be deployed in such coarser-grained ways, and indeed those are the variants I'm more sympathetic to.

Rather than syntactic tree-structures, we'll instead be working with (a generalized form of) directed graphs. And rather than representing coordinated thoughts with "wires" between multiple occurrences of Bob, we'll instead have Bob occurring just once in Alice's thought-graph (for each presentation Alice has of him). The directed edges will have to "come back" to that single occurrence each time they represent a coordinated thinking about him.

Thus, when Alice thinks the coordinated thought she does in case (3), we'll understand her to have thoughts with the following structure:

(7) 

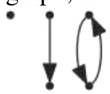
and when she thinks the uncoordinated thought she does in case (1), we'll instead have something like this:

(8) 

To develop this properly, though, we should first review some elementary graph theory.

4

A **graph** is made up of one or more **vertices** joined by zero or more **edges**. In some applications, the edges are assumed to be undirected; in others, not. Graphs where all the edges are directed are often called **digraphs**. Graphs may or may not be **connected**; here is an unconnected digraph, with three components:



and some vertices may be **isolated**, as the leftmost vertex in the above diagram. But usually *edges* are not allowed to be disconnected in an analogous way; they are usually assumed to always begin and end at specific vertices. In some generalizations, though, that we will consider later, this requirement on edges can be relaxed.

The edges in a graph may or may not exhibit **cycles**, as this digraph does:

But usually **self-loops** starting and ending at the same vertex are prohibited:

That's not because there's anything incoherent about self-loops; it's just that for many customary applications, it's easiest to assume these are excluded. Similarly, for many applications, it's easiest to assume there is at most one edge (in a given direction) joining a given pair of vertices. In the following diagram:

the edges on the right would violate this requirement. Again, this is not because there's anything incoherent about multiple edges joining a given pair of vertices: indeed, the famous Bridges of Königsberg problem, which launched the discipline of graph theory, makes essential use of such. It's just that for many customary applications, it's easiest to assume these are excluded. For our purposes, though, we will want to allow self-loops and multiple edges.[15]

Another requirement we'll want to relax is that edges must always join exactly two vertices. For some applications, one instead wants to work with structures like these:

Notice there is no vertex in the middle of the diagram. This graphlike structure has five vertices and a single edge that joins all of them. Graphlike structures of this sort are called **hypergraphs**. For these structures, we assume that each edge joins *one or more* vertices, rather than always exactly two.

The hypergraph in the preceding diagram had an undirected edge; we might instead designate some of the vertices it joins as initial and others as terminal, thus getting a directed hypergraph:

I want to refine these structures in a slightly different way. Instead of sorting an edge's vertices merely as initial or terminal, let's assign them a linear order.

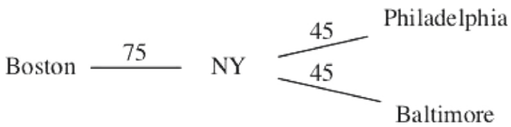
[15] Sometimes such structures are called "pseudographs" or "multi-graphs" rather than just graphs, but the literature is not consistent. A *directed* graph that permits self-loops and multiple edges is sometimes called a "quiver." State transition diagrams for formal automata are often of this sort. (They also have labelled edges, to be discussed below.)

Thus we will instead have something like this hypergraph, with an ordered length 4 *edge*:

That's to be distinguished from this simple digraph with a directed *path* of 4 edges of length 1:

Think of the latter structure as involving four different flight bookings; but the former as involving a single booking with three intermediate layovers. For our purposes, we will need the more general hypergraph structure that allows for both of these possibilities.

For some applications, we want to associate further information with a graph's vertices or edges. For example, in this (undirected) graph the labels on the edges may represent the cost of traveling between the cities that label the vertices:



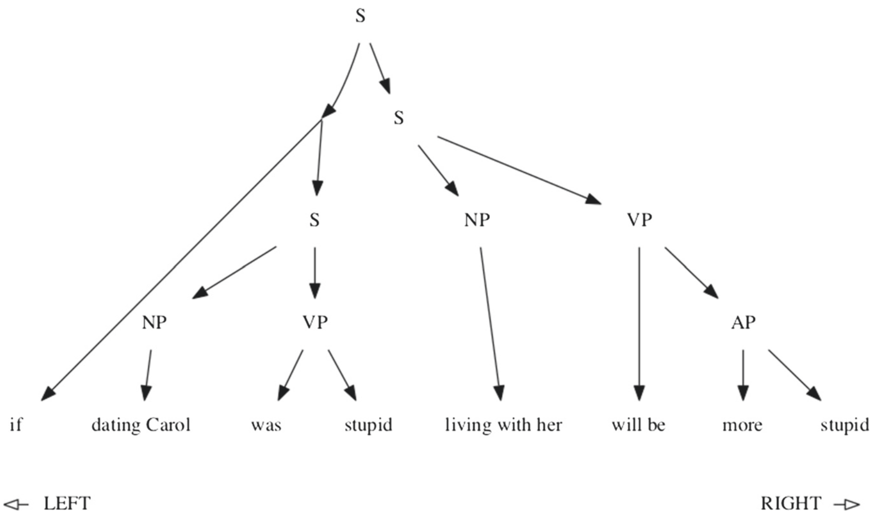
Notice we don't hesitate to assign one and the same label to distinct edges. What's trickier to understand is that we needn't hesitate to assign one and the same label to distinct vertices, either. This is because the information *labeling* the vertex shouldn't be thought of as *constituting* the vertex. Rather, we should just think of the vertex as constituted by *its position* in the graph structure.[16]

For example, consider the phrase-structure tree of a sentence where the same word appears multiple times:

(9) If dating Carol was *stupid*, living with her will be more *stupid*.

A phrase-structure tree can be thought of as a special kind of vertex-labelled digraph, with an additional left-to-right structure imposed on its leaves:

[16]In many expositions of graph theory, graphs are *constructed* using set-theoretic machinery from a pre-existing set of objects assumed to constitute the vertices, and an edge relation on them. With such constructions, the present point can be expressed by saying that there need be no relevant connection between the object a vertex is constructed from, and the information that vertex is labeled with. More significantly, though, the notion of a graph can be rigorously understood in other ways, too, that needn't identify specific pre-existing objects to constitute each vertex.



Notice that the single word-type "stupid" labels distinct vertices in this graph. (So too do the categories S, NP, and VP.)

Now, you might consider saying instead that it's the distinct *tokens* of "stupid" in our *inscription* of sentence (9) that should label the vertices of its phrase-structure tree. I don't know what account you'd then give of the syntax of (9)'s sentence-*type*. But this idea falters even as an account of the syntax of tokens. Consider the following inscription:

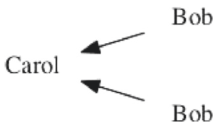
(9′) If dating Carol was
living with her will be more STUPID

This contains only a single concrete instance of the word "stupid," yet it too should have the same syntactic type as displayed above. But here we only have a single token of "stupid" to go around. So whether it be type or token, inevitably we'll have to admit that one and the same object sometimes labels distinct positions in a phrase-structure tree.

Some philosophers distinguish what I've called a token from what we call an *occurrence* of the word-type "stupid." Even if that word has only a single concrete tokening in (9′), they'd say it has two occurrences, one for each of the syntactic positions where the word appears. I have no complaint with that. But don't think that such occurrences are given to us as distinct objects in advance, out of which we could *build* the vertices of our phrase-structure tree. Rather, the distinctness of the occurrences instead already assumes the distinctness of those syntactic positions. And having been *given* the distinctness of the latter, there is no harm in letting it be the single word-type "stupid" that labels them both.[17]

[17]Richard 1990 p. 212 has an example like my (9′). Salmon 2006a, note 4 (and 2006b, note 1) also emphasizes the difference between tokens and occurrences.

Hence, labeled digraphs like this:

(10) 

should be clearly distinguished from labeled digraphs like this:

(11a) 

or this:

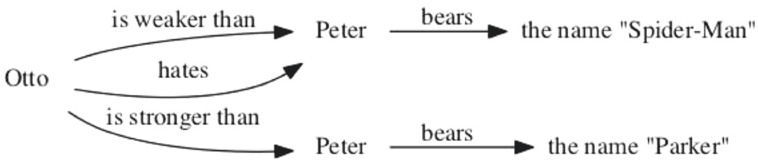
(11b) 

even when it's one and the same object, Bob, which labels both of the vertices on the right-hand side of diagram (10). The graphs are different because Bob appears twice (that is, he labels two vertices) in (10) where he only appears once in (11a) and in (11b).[18]

5

So how can we use these tools to model Alice's thoughts?

Let's suppose she turns her thoughts from baseball to the denizens of Marvel Comics. She doesn't realize that the individual named "Peter Parker" and the individual named "Spider-Man" are one and the same. Her thoughts may have a structure like this:

(12) 

This represents Alice as thinking *twice* about a single individual (here designated as Peter), who she thinks of in qualitatively different ways. As presented one way, the individual seems to her to be hated by Otto, and (only) to bear the name "Spider-Man"; as presented another way, he seems to her (only) to bear the name "Parker."

So far, we're only modeling thoughts whose content involves the atomic predication of binary relations. Since Alice thinks about Otto standing in multiple relations to Spider-Man, we need to allow her mental graph to join the upper-left vertices with

[18] A useful parallel is the notion of a **multi-set**, which is like a set in being insensitive to order, but like a sequence in being sensitive to multiplicity. The multi-set of vertex labels for graph (10) contains Carol once and Bob twice (in no order); the multi-set of vertex labels for graphs (11a) and (11b) only contains each of these labels once.

multiple labeled edges. We should also be prepared for her thoughts to join some vertices to themselves: for example, she may believe that Otto praises himself:

(13)
$$\begin{tikzcd} \text{Otto} \arrow[loop above, "\text{praises}"] \end{tikzcd}$$

Here then would be Alice thinking in a coordinated way about Peter, that both Otto and he will vote him Most Inadequate Hero:

(14)
$$\begin{tikzcd} \text{Otto} \arrow[r, "\text{will vote for}"] & \text{Peter} \arrow[loop above, "\text{will vote for}"] \arrow[r, "\text{bears}"] & \text{the name "Spider-Man"} \end{tikzcd}$$

By contrast, here would be Alice thinking about Peter in an uncoordinated way:

(15)
$$\begin{tikzcd} \text{Otto} \arrow[r, "\text{will vote for}"] & \text{Peter} \arrow[r, "\text{bears}"] & \text{the name "Spider-Man"} \\ & \text{Peter} \arrow[u, "\text{will vote for}"] \arrow[r, "\text{bears}"] & \text{the name "Parker"} \end{tikzcd}$$

This picture needs to be extended in various ways. But first let's talk about what these graphs are meant to capture. When representing Alice's thoughts, we might aim to model just her fully-committed beliefs, or perhaps those and her desires and intentions, or perhaps a wider range of propositional attitudes. This is what I do aim to do; we will discuss shortly how to bring more attitudes than belief into the framework. Additionally, one might aim to capture states that aren't propositional but are equally manifest in Alice's mental life, as some would construe states like seeking, imagining, and admiring, and others would construe phenomenal consciousness. Let's suppress all such states for the purposes of this discussion. Finally, one might also aim to capture lower-level facts about Alice's mental life. Perhaps one of her presentations of Peter comes quicker to mind when she considers who Otto interacts with. I am not here aiming to capture any facts like these, either. I'm assuming that for some explanatory purposes we want to model facts about Alice's thinking that she shares with other thinkers, where those subcontentful details may be different. That's the level of abstraction that these mental graphs are meant to work at.

Okay, so let's consider how to refine the graph framework towards that goal. One question is how to represent the difference between Alice's *believing* and her *intending* that Otto will defeat Spider-Man. The graph:

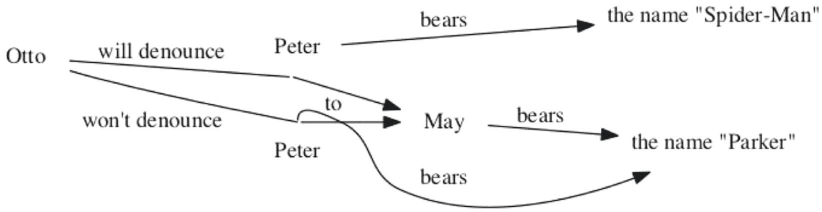
(16) Otto $\xrightarrow{\text{will defeat}}$ Peter

doesn't seem to distinguish these. Neither does it distinguish her beliefs from her *deliberate withholdings* of belief (which is not the same as a mere lack of belief and disbelief). A quick fix would be to label the edges not merely with the relation Alice's thought ascribes, but also with the *attitude* she ascribes it with; but we will suppress this issue for now and give a more considered response later.

A second question is how to represent non-atomic thought contents. Here too, we will give a more considered response a bit later. At present, let's just allow the possibility of Alice's thoughts ascribing (singly) negated relations:

(17) Otto $\xrightarrow{\text{won't defeat}}$ Peter

A third question is how to represent thought contents that ascribe *unary* properties, or that ascribe relations with arity greater than two. To handle the latter, we can help ourselves to the ordered hyperedges we saw in Section 4. For example, here is Alice thinking that Otto will denounce Spider-Man (but not Peter Parker) to May Parker:



(18)

Notice that it's a single (hyper)edge that leads from Otto, makes an intermediate layover at the upper vertex labeled Peter, then proceeds to May; and likewise a single edge that leads from Otto to the lower Peter to May. These two edges are labeled differently (only one is negated), because Alice thinks of Otto, Peter, and May as differently related when she thinks of Peter under each presentation.

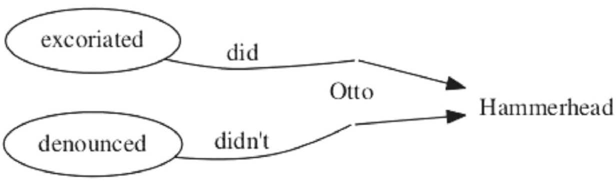
Similarly, unary predicates can be represented by labeled (hyper)edges that "join" only a single vertex: that is, which start there and point off to nowhere:

(19a) Otto $\xrightarrow{\text{is villainous}}$

Or perhaps we should draw it as:

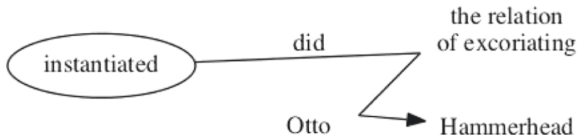
(19b) Otto $\xleftarrow{\text{is villainous}}$

A fourth question is how to represent Frege-type cases that concern the *predicative* parts of Alice's thinking, rather than the objects she's thinking of. One natural move here will shift the ascribed relation from its position labeling an edge into a new, distinguished kind of vertex. That is, suppose that the relations of denouncing and of excoriating are one and the same, but that Alice hasn't realized this. Then she might think that Otto excoriated Hammerhead while denying that he denounced him:



(20)

Notice that the (identical) relations of excoriating and denouncing here occupy specially marked vertices. We are not conflating Alice's thoughts with the second-order thoughts that Otto and Hammerhead did and didn't *instantiate* the relations of excoriating and denouncing. Those would be represented differently:



(21)

Here it's the higher-order relation of instantiating that occupies a special vertex, whereas Alice thinks of the relation of excoriating as just another object. This quick sketch is all I'll give to indicate how the framework might be extended to model Frege-type cases for the predicative parts of thought; for the remainder, I'll suppress such issues.

A fifth question is how to represent Alice's thoughts when they're partly "empty," that is, some of their expected objects are missing. A quick fix would be to represent her thoughts in those cases as joining vertices that lack any label. However, we will spend a little time on this and instead give a more systematic proposal, that can be repurposed when we address the task of modeling Alice's quantificational thoughts.

6

The tool we need for this is the notion of a **disjoint sum** or union. This is a counterpart to the more familiar notion of a Cartesian product. The product of two sets Γ and Δ is a collection of objects each of which designates *both* an element from Γ *and* an element from Δ. The sum of Γ and Δ, on the other hand, will be a collection of objects each of which designates *either* an element from Γ *or* an element from Δ. Initially, you might think this role could be played just by the ordinary *set-theoretic union* of Γ and Δ. But recall that when Γ and Δ overlap, we say that (u, v) and (v, u) may be *different* members of their product. In somewhat an analogous way, it will be useful for the counterpart notion of a sum to distinguish between a member that designates *u qua* element of Γ and a member that designates *u qua* element of Δ. So we define a disjoint sum not as a simple union, but instead as a collection of pairs, whose first element is a tag or index indicating which of the operand sets (Γ or Δ) the second element is taken from:

$$\Gamma \uplus \Delta = (\{\text{the tag } \textsc{left}\} \times \Gamma) \cup (\{\text{the tag } \textsc{right}\} \times \Delta)$$

It does not matter what objects we use as our tags LEFT and RIGHT, so long as they are distinct. These objects may even belong to the sets Γ or Δ; no confusion need result.[19]

[19]The disjoint sum operation is sometimes represented using the symbols $\uplus$, or $\sqcup$, or $\amalg$, or even $+$. In some definitions, the operand sets are themselves used as tags; but this isn't adequate for the case of taking the disjoint sum of a set with itself, which some applications do want to allow.

The notion so defined is a basic building block in the type theories of functional programming languages like Haskell and OCaml. There, instead of:

(22a) (LEFT, u), or
(22b) (RIGHT, u),

one would instead write:

(23a) Left u, or
(23b) Right u.

In the special case where the lefthand operand is a singleton set containing a special object supposed to represent a FAILURE, Haskell writes:

(24a) Nothing, or
(24b) Just u

instead of:

(25a) Left FAILURE, or
(25b) Right u.

Similarly, OCaml writes None or Some u. I will follow the Haskell convention.

How can we apply this to the case of Alice's empty thoughts? Let's suppose that, though Spider-Man and his kin *do* exist, non-Marvel-approved substances like kryptonite and phlogiston do not. But Alice doesn't know this. She thinks, confusedly, that Otto hoards kryptonite. She does not think that he hoards phlogiston. At least, that's how we'd like to put it. Can we represent this difference in Alice's thoughts, without any objects such as kryptonite or phlogiston to label her mental vertices?

What we will do is label Alice's vertices using not the bare objects of her thought Otto, Peter, and so on — elements of the domain **Marvel** — but instead using members of the disjoint sum {FAILURE} $\uplus$ **Marvel**:[20]

(26)
$$\begin{tikzcd} & \text{Nothing} \arrow[r, "\text{bears}"] & \text{Just the name "phlogiston"} \\ \text{Just Otto} \arrow[r, "\text{hoards}"] & \text{Nothing} \arrow[r, "\text{bears}"] & \text{Just the name "kryptonite"} \end{tikzcd}$$

Notice that the labels of these vertices are no longer the objects of Alice's thought, even when there is such an object. Alice is not thinking about the abstract construction Just Otto. She's thinking about Otto. Neither is she thinking he hoards the abstract construction Nothing. There *isn't* any object she's genuinely thinking is hoarded, when she has the thought she'd articulate as:

(27) Otto hoards kryptonite.

[20]This strategy is suggested in Kamp 1984/85, p. 252, and Kaplan 1989, note 23 on "Plexy". But we execute the idea using the systematic framework of an "option type," as expressed in (24ab) or (25ab), rather than in the *ad hoc* way Kamp and Kaplan propose.

Neither is there any object for her to genuinely think about and call "phlogiston." But we haven't here represented her as also thinking:

(28) Otto hoards phlogiston.

(Neither does (26) represent her as thinking Otto *doesn't* hoard phlogiston, but a different mental graph could do so.)

If we want to represent "empty" thoughts in this way, then we'll need to go back and relabel all our graphs (12)–(21) using labels like the abstract construction Just Otto instead of simply Otto. Let's take this as having been done.

Another refinement will be to use as labels not members of the disjoint sum $\{\text{FAILURE}\} \uplus$ **Marvel**, but instead members of the slightly different sum $\{\text{FAILURE}, \text{VARIABLE}\} \uplus$ **Marvel**. The role of this special object VARIABLE will be to represent not thoughts that lack objects but bound argument positions in quantified thoughts, which we will discuss below. Before we get there, let me help myself to an extension of the convention by which we express (25ab) as (24ab). Instead of:

(25a) Left FAILURE, or
(25b) Right u, or
(25c) Left VARIABLE,

I will now write:

(24a) Nothing, or
(24b) Just u, or
(24c) Var.

That concludes the basic pieces of the framework I want to sketch. What remains is to discuss:

- different ways the framework can be extended to handle non-atomic thoughts (including quantified thoughts)
- the differences between this framework and neighboring views, such as neo-Fregean or mental file models of thought
- how this framework can be put to good use, for example in giving a semantics of folk attitude ascriptions

Sections 7–9 are devoted to the first two topics. I leave the third, ambitious, topic for later work. Section 10 summarizes what I hope to have achieved in this discussion.

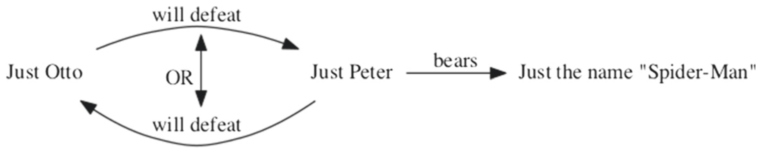
7

Let's start now to expand these models to logically complex thoughts. We can't just model disjunctive thoughts by *sets of graphs*, because that would introduce indeterminacy about which vertices in one graph purport to be about the same objects as vertices in another. Our account needs to be more integrated than that.[21]

[21] For some purposes, though, such indeterminacy would be an appropriate part of the model. In what follows, I am in effect assuming that in all of Alice's "attitudes worlds," the same objects exist and there

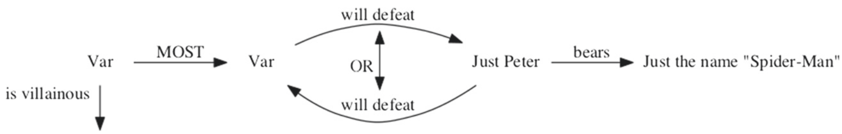
As I said at the start of Section 3, what I'm offering is deliberately incomplete. How we should continue here depends on our explanatory needs. In some settings we may want to model Alice's thinking in a fine-grained way. Perhaps not every syntactic detail from how she articulates her thoughts should be mirrored in the thoughts themselves; but we may want a model that tends in that direction.

If that's what we want, then we can extend the framework sketched so far as follows. First, connectives like or and and can be modeled by introducing labeled edges that join not vertices but other edges.[22] Here might be Alice's thought that one of Otto or Spider-Man will eventually defeat the other:



(29)

Using the tools introduced in Section 6, we could model the quantificational thought that most villains will either defeat or be defeated by Spider-Man like this:



(30)

This quick sketch should give a feel for the idea.[23] Rather than developing it further, let's instead think about other ways to extend the framework, that are more coarse-grained and encode less language-like structure into Alice's thinking — though they do still preserve the difference between coordinated and uncoordinated thinking. In some settings, this may suit our explanatory needs better.

are no open questions about their cross-world identification. This is of course an unrealistic idealization. In a fuller account, we'd need to introduce more looseness into the system, so that questions of "trans-world identity" aren't always so cut-and-dried. The models presented here would be natural parts of a more complex representation.

If the graphs I describe are made maximally opinionated, then they can serve as epistemically possible worlds and what glues them together would be epistemic counterpart relations, which needn't be transitive or one-one. But I'm not primarily thinking of mental graphs as maximal in that way.

[22] Chris Barker observed that such structures are also used in Johnson and Postal 1980. Dan Hoek observed that these graphs, unlike the previous ones, no longer have the property that all their subgraphs also partly model Alice's thinking.

[23] The direction I'm going here is close to the proposals by some discourse representation theorists to use their apparatus to model attitude states as well as natural language meanings. See Kamp 1984/85 and 1990, and Asher 1986, 1987, 1989. I am broadly sympathetic to those proposals — though they tend to be more descriptivist in their handling of "internal anchors" than I would be, and I take very seriously Kamp's own admonition that "we should not take it for granted that *every* feature which is needed in a semantic theory such as DRT must necessarily have its psychological counterpart" (1990, p. 42). Still, the framework presented here and those developed by Kamp and Asher draw on many of the same guiding intuitions and are at least spiritual cousins. I haven't adopted their theoretical apparatus more closely, because (i) I prefer Amsterdam-style machinery for working with natural language dynamic meanings; and (ii) I seek an account that's less committed to particular pieces of syntax ("reference markers") or private mental representations than DRT is, on its natural interpretation. We'll return to this last point in notes 31 and 39 and in Section 10, below.

Let's suppose the coarseness of grain we're aiming for is to identify all of Alice's thoughts (under a given attitude) having the same prenex/disjunctive normal form. Thus, the thought she'd articulate as:

(31) If there is a villain who likes pumpkins, many children will be delighted but Spider-Man will not.

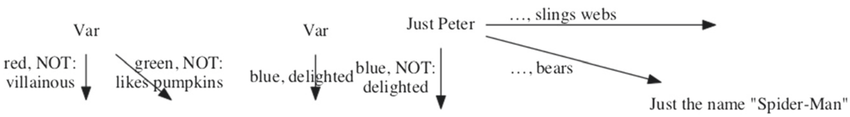
might be equivalent to:

(32) $\forall x$. Many c. ($\neg$ Villainous x) $\vee$ ($\neg$ Likes pumpkins x) $\vee$ (Child c $\wedge$ Delighted c $\wedge$ $\neg$Delighted Spider-Man)

For this variation, rewind to graphs as described in Sections 5 and 6. So the edges only represent Alice's atomic thought contents, except we're also allowing them to represent (single) negations of such. Each (hyper)edge joining n vertices has a direction or order for those vertices, as well as what we might call a "polarity," indicating whether the n-ary relation it's labeled with is thought by Alice to hold or to not hold. Let's extend this by also supposing each edge to have one or more "colors." My idea here is to impose a **cover** on all the edges in Alice's mental graph. A cover of E is a collection of non-empty subsets of E whose union exhausts E; that is, it is like a partition except that cells in the cover are permitted to overlap. The metaphor of "coloring" the edges is just a vivid way to think of this. Each "color" had by an edge represents an additional cover-cell to which the edge belongs. (We don't rely on any other facts about the "colors" besides their identity and distinctness.) The point of these cover-cells will emerge in a moment.

If you look at the form of Alice's thought (32), you'll see that it involves a sequence of quantifiers, around a disjunction each of whose disjuncts is a conjunction (of one or more terms). Each conjunction can be identified with a set of atomic (or singly negated) predications. We will associate each such conjunction with a new color. (If a given conjunction recurs in other thoughts, the same color should be re-used.) Then the whole disjunction can be identified with *the set* of those colors.[24]

Hence, the quantifiers aside, Alice's thought (32) might be modeled like this:

(33)


One of the Var vertices corresponds to the bound variable x in (32), and the other to the bound variable c. The complex disjunction which is inside the quantifiers in (32) can be represented by the graph displayed in (33), together with the set of colors {red, green, blue}. To get Alice's whole thought, we have to also specify the surrounding quantifiers. This can be represented by an ordered sequence of zero or more pairs, each pair consisting of a quantifier and some Var-labeled vertex in the graph. And we should also specify *the attitude under which Alice thinks* the thought in question: does she believe it? intend for it to be true? remember it? or what?

[24] We are ignoring differences between $q \vee (r \wedge s)$, $(s \wedge r) \vee q$, and so on. This is in keeping with the coarse-grained aims of the present approach.

Altogether, then, we can define an **attitude triplet** to be a triple of an attitude, a sequence of zero or more (quantifier, Var-labeled vertex) pairs, and a set of colors.[25] Given a graph as in (33), each such triplet will represent one of Alice's thoughts; and a set of them can represent all her thoughts. Coordination relations can be identified between her thoughts, as well as within them, because those thoughts will involve edges joining a common vertex in the underlying graph. Again, this is just a quick sketch, but it should give enough of a feel for the idea.

I'm deliberately not advocating one of these variations over the other, because I think Alice's thinking will plausibly have *each* of the structures described: both the fine-grained one and the coarser-grained one — and presumably others too. Different theories may want to make use of different structural properties of her thought. What unites the different models I've sketched here is their shared strategy of representing the difference between coordinated and uncoordinated thinking in terms of how many mental vertices are present in the underlying graph, as opposed to the "wires" and "unwires" from the start of Section 3.[26]

8

The framework I've presented *looks* a bit different from the "mental file" and neo-Fregean theories we're more familiar with.[27] But perhaps these differences are merely superficial?

As I mentioned in Section 1, the machinery developed here needn't be understood as competing with a "mental files" theory. The most prominent difference is that in my framework, Alice's memories, information, and attitudes are *distributed across* thought structures, that may involve multiple "files" or vertices, rather than being *contained inside* each file. This gives a more satisfying representation of thoughts that predicate ($n > 1$)-ary relations.[28] If we had to bundle Alice's relational thoughts about Otto into a single file, how should that bundled information latch onto her presentation of Peter as Spider-Man rather than his presentation as Peter Parker? Minimally, we'd have to take care to distinguish her thoughts about Otto's relation to *the person Peter Parker*, as presented by some other file, from her thoughts about *that*

[25] Kamp 1990, p. 46 also uses the expository device of "coloring" parts of the structures he's working with; but in his deployment, the color represents whether that part of the structure models a belief, or a desire, or so on. His official proposal separates the attitude out in the same way my triplets do.

[26] For a possible complication, see note 43, below.

Examples like "Carol came and so did Daniel; *one of them* left their coat" (compare example 4 in Heim 1983) show that the number of vertices don't correspond to the number of objects the subjects thinks are (or even may be) present. Landman 1990, pp. 273ff has an interesting discussion of vertices when it's epistemically unsettled how many objects they stand for. See also van Eijck 2006, Section 7.

[27] See Recanati 2012 and some of the precedents he cites on p. vii. Compare also the use of "discourse referents" or "reference markers" or "pegs" in formal semantics, in discussions like Kamp 1981 or the papers cited in note 23, above. Heim's dissertation and her 1983 are commonly cited as early contributions to each of these traditions.

[28] Mental file theorists often give this minimal attention. See for example Recanati 2012, Chapter. 4.4 and 15.1, and Perry 1980, pp. 73–4.

mental file or presentation.[29] More awkward would be questions why Alice's Otto file is entitled to populate itself with information that makes use of other mental files in a way that's itself *immune* to reference-failure and Frege mistakes.[30] (If mental files have access to such a representational system, why can't we?) These issues are not insuperable for a mental file theorist, but I expect that satisfactory answers to them are just going to reproduce the structure of the framework I've presented. They may be willing to take this on as a development of their initial account. If so, I have no complaint.

If my framework *were* construed as a variant of a mental files account, it would be one that allows files with no referent, files that are not even *believed* to have a referent (like the Var-labeled files in (30) and (33)), and the possibility of distinct but indiscernible files. Recanati 2012 explicitly allows the first two of these. He doesn't address the third, and some of his arguments trend against it, but on the whole it seems to me compatible with his fundamental picture.[31]

In Section 1, I also cited Ninan's "multi-centered" view of mental content. The graph models sketched above use different theoretical apparatus than Ninan, but I think our proposals are fundamentally similar. Ninan needs an arbitrary ordering of the objects of an agent's thought, which we might avoid (though the most common mathematical constructions of graphs don't bother to be very hygienic about this, see note 16, above). Also, Ninan's proposals about mental content are intertwined with his proposals about the semantics of attitude reports; and here I want to diverge from

[29]Some discussions like Heim 1983 are sloppy about this, but we can use the machinery introduced in Section 6 above to be more careful.

[30]In a Unix-type filesystem, a file is individuated by what device it resides on and its "inode number." (*Pathnames* can't individuate files, because they can stand many : one to inodes.) In general, an application or even the whole OS may not be in a position to tell whether "two" devices host distinct filesystems, or are merely a single filesystem being accessed via two routes. So even to a whole computer, numerical identities between files need not be luminous or transparent.

[31]Of relevance here is Recanati's contrast between files and the information they contain; see pp. 38–9 and 80–82. A key question is whether he'd permit Ellen in the case I discuss below to have two mental files.

See also Recanati's pp. 244–7, where he likens mental files to representational vehicles of the sort Fodor appeals to. This aspect of Recanati's view does pull apart from the framework I'm advocating. In general, much of the "mental file" literature, as well as the semantics literature appealing to "discourse referents," can be interpreted in more "symbolist" or in more graph-theoretical ways. The "permutation" worry I articulate in Section 9 below is a useful diagnostic: to the extent that a view countenances real permutations, it's more symbolist. Symbolist versions of "mental file" theories (like Crimmins, and Recanati in the passages cited) tend to sound like Fodor and the other examples from note 1 above; symbolist versions of "discourse referent" theories (like Kamp and Reyle 1993, pp. 61ff) usually talk about syntactic objects like variables or numerical indices. Much of these literatures is not clear and explicit whether it means to be understood in the symbolist way or more abstractly; and even when authors are explicit, they tend to use the same vocabulary and cite as precedents other authors who fall elsewhere on the spectrum. For example, Asher 1986 uses the term "peg" to mean a variable or a private mental representation, whereas Landman 1986 and 1990, pp. 276ff, insists on using the same term to mean something extra-syntactic and intersubjective, more-or-less equivalent to our graph vertices. Landman writes: "Pegs are not mental objects... Pegs are postulated objects, whose function is to keep track of objects in information states" (1990, p. 277; he goes on to say that the same pegs exist across different conversations and even languages, compare Cumming 2014). I will presume to interpret all these views in the more abstract, less symbolist way.

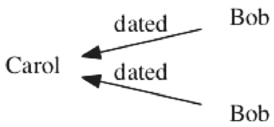
him more substantially. But I cannot adequately address the semantics of reports in this paper.[32]

Neo-Fregeans can also agree with some parts of my framework. Or at least, my framework seems capable of representing things they want to say. Information that Alice associates with Peter, when presented in a given way, is represented in her graph by the edges incident with that Peter-labeled vertex. For example, the presentation of Peter deployed in her thought (31)/(32) is connected with the information that he is named "Spider-Man" and that he slings webs — even though this additional information isn't itself part of that thought. (That's why those edges in graph (33) weren't colored red, green, or blue.) Nor are Alice's associations limited to beliefs: the edge that associates Peter with web-slinging could be part of a suspicion, or an intention, or some other attitude. So this framework seems hospitable to the neo-Fregean's ambition to capture the informational and attitudinal differences between Alice's different presentations of Peter.

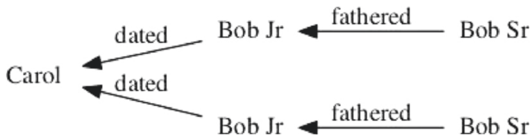
Still, I count it as an *advantage* of my framework that is does not *force* us to say that Alice associates rich bodies of information with every mental vertex. Her presentations can manage to be multiple even when the information she associates with them is comparatively thin, and in particular falls short of being uniquely identifying.

Indeed, her vertices or presentations can be multiple even when they're associated with *all the same* information and attitudes. As I suggested at the outset, Alice's two presentations of Bob in case (1) might differ merely numerically, as do the vertices on the right-hand sides of diagrams (10′) and (34):

(10′)



(34)



[32]Ninan also associates his multiple centers with guises or modes of presentation. He may have several motivations for this, but none of them seem forceful in our dialectical context. One is the connection between his account of mental content and an account of the semantics of attributions; as I said, I'll be offering a different account of attributions elsewhere. Another is to provide a valid domain for Ninan's "tagging function"; but this could also be accomplished by having that function map *particular centers* (positions in the sequence of objects), rather than pairs of objects and guises, to epistemic counterparts. Lastly, if a set of multi-centered contents is meant to provide only a *partial* characterization of what attitudes an agent has, including guises may provide some information about other attitudes whose content is not specified. On the other hand, if we're discussing *total* models of an agent's attitudes (which is not the same as being "maximally opinionated", in the sense of note 21, above), I don't see any need here for guises. Cognitive differences should either be *represented in* the totality of what the agent thinks, or else they can be ignored as irrelevant to the explanatory aims pursued here. (Remember, I allowed in Section 5 that there may be interesting cognitive differences my models exclude.) I'm not interested in encoding information about *the means by which* an agent thinks of various objects that goes beyond differences in her total attitudes towards them.

So, though my framework can *accommodate* all the associations and fineness of grain a neo-Fregean wants, adopting it may undermine, or at least relieve the motivational pressure for, paradigmatic neo-Fregean claims about *how much* associated information must be available.[33]

Should we *want* our models of Alice's thinking to have the capacity to represent merely numerical differences in her presentation of a single object? Is it really intelligible for her to have multiple presentations of Bob that are "qualitatively indiscernible" — that is, unaccompanied by any qualitative difference in her associated memories, information, or attitudes? (As in (34), we allow there to be networks of *further numerical* differences in what other, also qualitatively indiscernible, vertices they are adjacent to. The upper Bob Jr vertex is joined only to the upper, and not also to the lower, Bob Sr vertex.)

Here are several cases to motivate the intelligibility of this.

Suppose Alice has a vivid dream of a distinctive-looking woman, and remembers the experience. The dream did not present itself to her as being of anyone she was already familiar with. We can suppose also that it did not present itself to her as being dreamt, or at any rate that over time she forgets that this experience came from a dream. Later in life she has a perceptual experience that exactly qualitatively matches the dream experience. Most will agree that it's possible for the dream not to have been a prescient presentation *of* the same person she later perceives.[34] Intuitively, neither do these two experiences need to be coordinated. It can be an open epistemic question for Alice whether these two presentations were of a single individual. Twisting the case, we can further suppose that the later experience was in fact itself a hallucination. That doesn't seem to render the two presentations any less distinct. We can suppose further that as time passes, Alice forgets which experience came first, whether one was dreamt and the other apparently perceptual, or what. They might nonetheless remain distinct presentations to her. Yet no merely qualitative information contained in the presentations may differentiate them. For a final twist, we can suppose that Ellen is a mental duplicate of Alice as she is now, but in Ellen's case, the two memories of a distinctive-looking woman never were generated from distinct experiences. She only ever had a *single* experience, but through some signaling glitch in her brain, now has cognitively distinct memory presentations, just like Alice. If Alice or Ellen were to revise one of their memories, it would change autonomously. The other presentation would remain as it is.

Suppose Flugh is an alien whose many eyes are on long stalks, which can wriggle through the maze of twisty little passages where he lives. The experiences generated by his eyes are co-conscious, but not always spatially integrated. In particular, on one day he has two qualitatively matching experiences of a homogeneous sphere, without any presentation of how the spheres are spatially related to each other. Neither experience even *seems* to be above, or to the right, of the other. As it turns out,

[33] As Heck 2012 observes (esp. pp. 150–59), the critical Fregean contribution to many explanations is just *patterns of identity and difference* in sense.

[34] Compare Kaplan 1968, pp. 198–9; Kripke's discussion of unicorns and Holmes in his 1980, Addendum A; and Kamp 1990, pp. 59–61.

Flugh is seeing only a single sphere.[35] In fact, it may be that he's only seeing a single sphere with a single eye, but through some signaling glitch in his brain, he now has cognitively distinct mental presentations. That is, Flugh is in a position to wonder:

(35) Is *this* sphere the same as *that* one?

Perhaps Flugh would be in a position to *introduce* some difference in how he thinks of the two spheres. Perhaps he could start thinking of one of them as the sphere I'm attending to in memory right now. That may make his thought problematically private and ephemeral, but regardless, I don't see why Flugh *must* be thinking of the spheres that way. Couldn't his two sphere presentations *in fact* be qualitatively exactly alike, but nonetheless be *two* presentations?[36]

Also relevant is some interesting empirical work by Anthony Marcel. Following some informal reports in Schilder 1935, Marcel established that retinal disparity and splitting attention between vision and touch could generate an illusory tactile experience as of being touched twice on one's fingertip, when in fact there was only a single touch. Interestingly, subjects were confused about the felt location, or even the spatial relation, of these two (apparent) touches. Marcel describes the effect as "perception of tactile two-ness without determinate perception of each of the two touches" (p.c.). The phenomenon needs to be investigated further, but one intelligible possibility is that the touches and their felt locations were presented as two, without there being any difference in the qualities they were presented as having. Neither touch felt to be to the right of the other; subjects were not even sure on what axis the touches were felt to be aligned. That's why identifying and comparing the felt locations was so difficult. Marcel does say that when he underwent the experiment himself, he "came to be able to perceive one touch as lighter than the other but was unsure which was which." But that leaves open the possibility that before he achieved this, or in the experience of other subjects, there was no qualitative difference between the two touch experiences.

In response to such cases, some of my interlocutors protest that the subjects would only be thinking *plurally* about *a pair of (putative) women*, or *a pair* of spheres or touches, and wouldn't be in a position anymore to think about *each* element of

[35] This kind of case has been reinvented many times. I first constructed a version of it while a grad student, to combat the idea that co-conscious perceptual experiences always had to have a single perceptual viewpoint. But I was anticipated by David Austin's "two tubes" example in his 1990; and as Austin notes, he was himself anticipated by Gettier and Rosenkrantz. See also Austin's case of Sarah the astropharmacist (pp. 42ff), and Fine's "two Bruces" in his 2007, pp. 36–7, 42, 70–71. I prefer my alien example to Austin's human one, because it's essential for the subject to have two disjoint visual fields, with no spatial relation to each other (see Austin pp. 20, 45 note 15, 108 note 1, 124). It's hard to imagine a (contemporary, neuro-surgically unaltered) human's phenomenology being like that. If you try Austin's experiment yourself, you'll find that *your* visual experiences still *are* presented as spatially related: one of the apparent spots will look to be to the right of the other.

Austin's case has been discussed by many authors, including Levine 1988, who invokes token mental symbols to explain the psychological difference (compare notes 1 and 5, above); Boër 2007, pp. 43–4, 253ff adopts a similar strategy. See also David Sosa's 2010 response to Fine's "two Bruces", at pp. 350–51, 355. Austin is also cited repeatedly by Chalmers (see the citations in note 3, above). Chalmers takes the example to motivate additional "centers" or demonstrative anchors for a qualitative "scrutability base."

[36] Here is where it's especially natural to invoke multi-centered contents, as Ninan does.

the pair separately.[37] I acknowledge there can be cases of that sort. But I don't see why our cases must take that form. This diagnosis sounds especially forced for the dreamt woman case, where the two presentations never need to have occupied Alice's attention simultaneously.

Others may insist that in my examples there must be some underlying difference in how Alice's or Flugh's twin vertices are *implemented*. That may be; but at the level of abstraction we're modeling their mental reality, which can be shared between them and other physiologically different subjects, I don't see why any such implementation differences need to be manifest.

Consider also this. Flugh may know that spheres have invisible properties, let's call them *even-centeredness* and *odd-centeredness*, which are roughly equally distributed. If he's confident that the spheres he seemed to see are distinct, then it seems that two lines of hypothetical reasoning should be open to him. One would go: Suppose *at least one* of the spheres is even-centered, then the probability that they're both even is $\frac{1}{3}$, because in $\frac{1}{3}$ of the cases where at least one of two independent spheres is even-centered, they both are. But there's also another line of reasoning available to him. It would go: Call one of the spheres X. (Perhaps here he calls to mind and attends to one of his sphere images; or perhaps he doesn't.) Suppose *it* is even-centered. Then the probability that they're both even is $\frac{1}{2}$ — again because the centeredness of the other sphere is an independent matter. Both of these seem to be legitimate courses of reasoning. They are marred only by Flugh's ignorance of the fact (and giving no credence to the possibility that) the two spheres are one and the same. It may be a subtle matter for Flugh to sort out when he should deploy the one line of reasoning and when the other. But even making sense of the *intelligibility* of the second line of reasoning suggests that Flugh will have, at least temporarily and suppositionally, the ability to think about his "two" spheres independently.[38]

9

Some questions and replies.

First, suppose neo-Fregeans are prepared to accept the possibility of merely numerical differences in Alice's presentations of Bob. How much of a departure need this be from more familiar versions of their view?

Some neo-Fregeans will already insist on the possibility that modes of presentation be "non-descriptive." Often what this seems to mean is only that the subject may lack *concepts* or the *language to articulate* information that her presentations can nonetheless ascribe to an object. I haven't made any effort to exclude or include this in my framework. If we did want to specially keep track of such "non-descriptive" modes of presentation, though, they'd presumably have to do with what labels *the*

[37] Recall our opening remarks about example (1) versus (2). Evans pushes in this direction when he discusses indiscernible spheres in 1982, Chapter. 4.

[38] Thanks to Dan Hoek for encouraging me to press this. See also Stalnaker 2012; and Halpern 2003, pp. 1–2, 82–3, 213–17. Also relevant is the linguistics literature on "indistinguishable participants," in sentences like Kamp's "If a bishop meets another bishop, he blesses him." These were first discussed in Heim 1990.

edges of Alice's graph, not with the number of indiscernible vertices she can have. As I'd put it, this is only an issue about *which* qualities Alice's presentations can attribute, or perhaps the formats in which she grasps those qualities. What I'm after is more radical: the possibility of presentational differences that exceed *any* differences in attributed qualities.

Some neo-Fregeans like Evans will also already insist on the possibility that presentations may be qualitatively indiscernible but distinct, because they are presentations of different objects — or at least, presentations grounded in "different ways of gaining information" from an object. This is a step in the direction I'm after. But I want to go further in that direction than I expect Evans would. In some variants of the examples in the previous section, Ellen and Flugh only gained information in a single "way," but due to signaling glitches that information was cognitively "forked," and they then had multiple presentations downstream.

Let's suppose the neo-Fregeans are willing to accommodate all of this. All my examples motivate, they may say, is the possibility that senses may differ *solo numero*. This may be a relatively novel variety of neo-Fregeanism, but why shouldn't it belong to their family of views for all that?

Indeed, that would grant most of what I'm after. If you think of yourself as a neo-Fregean, and you can understand the proposals I'm making as refinements of your own view, I see no compelling need to resist that. The only reservation I have about this way of framing my proposals is this. If Alice really does have two numerically distinct but qualitatively indiscernible presentations of Bob, and these presentations are reasonably thought of as two *particular senses*, then mightn't it make sense to talk about *two* (or more) possible situations, in which these particular senses are permuted. I don't think that *ought* to make sense. There doesn't have to be anything in the mental reality we're modeling — at the level of abstraction we're modeling it, which may be shared between Alice and others who realize the same mental structure in concretely different ways — that captures such differences. At this level of abstraction, the permutations aren't meaningful.[39] Properly understood, that's the result a graph framework for talking about Alice's mind should deliver. There is no graph-theoretical difference between our graph (10′) and a "different" graph in which the two right-hand vertices have been permuted. If we talk about *particular senses*, though, I worry that these permutational differences might be genuine possibilities. At least, I think we'd have to do work to earn the right to say that permuting the senses makes no difference. If we can achieve an understanding of indiscernible but numerically different senses where such permutations *aren't* genuinely different possibilities, then okay, we're just doing graph theory in other vocabulary.

Second, in presenting this framework and sketching different ways it may be fleshed out, I've only been specifying what it is for an agent to have a given *set* of attitudes *at a single time*. I haven't said anything about what it would be for the agent to retain some of those attitudes while revising others. Nor have I said what it would

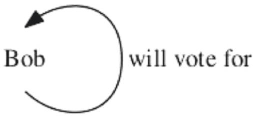
[39]This is an issue facing some linguistic accounts of "discourse referents" or "pegs." See, for example, Heim 1983, pp. 233, 242–3; and Groenendijk et al. 1996, pp. 194–95. See also the discussion of Ninan needing an "arbitrary ordering," in Section 8, above.

be for the agent to share some of those attitudes, but not others, with other agents. I point this out to counter the impression that the framework developed here must give *holistic* answers to those questions: that the only kind of mental overlap it would allow would require sharing entire mental graphs. Not so. This framework can be developed that way; but it is also compatible with various stories about when agents share more local, specific mental structures.

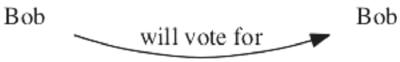
One especially hard version of the preceding questions concerns the mental relations between Alice, who has two presentations of Bob (and Peter and other objects), and another subject Daniel who has only a single presentation. When Daniel thinks about Alice's thoughts, should he associate her Bob_i presentation with his own unitary presentation of Bob? What then should he associate her *other*, Bob_j presentation with? These are complex issues[40]; and I mention them here only to disavow any ambition for the present framework *on its own* to be adequate for answering them. An answer will have to come from a fuller story about the semantics of our talk and thought about others' minds. I do believe the present framework is a useful component to be used in such a fuller story, but that will have to be shown elsewhere.

For the time being, then, I'm only offering a model of a single agent's mind, at a single time, concerning non-semantic and non-intentional matters.

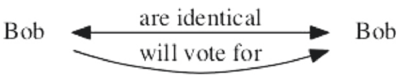
Third, I said earlier that coordinated thinking should be represented with a common vertex, like this:

(7) 

and uncoordinated thinking with multiple vertices, like this:

(8) 

But what if the agent has multiple vertices but *believes* the objects so presented to be identical? That is, what is the difference between (7) and:

(8′) 

(Here we meet again some "good questions" we raised in Section 3, but never answered.)

I expect that subjects *can* have attitudes like that, attributing identity to objects thought of in uncoordinated ways.[41] So I'll agree that (8′) represents a possible

[40]See the discussion of intersubjective transitivity in Richard 1990, pp. 210ff; Taschek 1998, pp. 347ff; and Fine 2007, pp. 107ff. See also Recanati's "vertical links" between one's own mental files and those of other subjects one is mentally representing (2012, Chapter. 14–15). Kamp's "internal links" (1990, pp. 74ff) and Asher's "internal connections" (1986, 1989) play the same role. Also relevant here are the symmetry violations discussed in Edelberg 1986 Section 3 and 1992 Sections 9–10.

[41]Various mental file theorists also make the point that one can come to believe identity holds across two files, without thereby merging the files. See Crimmins and Perry 1989, p. 692; Forbes 1990, p. 543; Recanati 2012, Chapter. 4, 8.3, and 15.2.

mental state, distinct from (7). Depending on *what attitude* it represents the subject as having, it may even be a *rational* state — for example, if the agent isn't fully, indefeasibly certain that the identity holds. If she *were* that fully, indefeasibly certain that the identity held, but persisted in thinking of the objects with two vertices rather than one, then perhaps she'd thereby be manifesting some irrationality.[42] So: (8′) might represent even a possible state of maximally-committed belief, distinct from (7), though arguably not an "ideally rational" one.[43]

There can also occur cases where the agent *does* aim to think of an object coordinately, but has unknowingly latched onto multiple referential candidates. In such cases, I'd posit a single vertex, labeled in some special way to indicate that the referent is indeterminate between the several candidates. (Here we can again use the machinery introduced in Section 6.) That would need to be clearly separated from vertices with *plural* referents, which we will also sometimes want. The framework sketched here naturally lends itself to such amplifications. The difficulties hereabouts concern what theoretical constraints and predictions we want to make with these various notions, not how to adjust the formal model to express them.

10

What I've presented here is only a programmatic framework; and as I've said, it may be developed further in different ways.

I observed some commonality between my framework and views advanced by Heck, Ninan, and Recanati, or at least with some core parts of their proposals. The divergences between us are interesting but are compatible with a shared fundamental spirit.

[42] Though, as I argue in other work, that's not to say that a *more* rational state will always be available to her. Also, Ned Block observed that it may not be under the subject's rational control whether her mind organizes itself in way (7) or way (8′).

[43] On the other hand, perhaps it *may* sometimes make rational sense to knowingly think of one object via two certainly co-referential mental presentations. If we decide to countenance this, we should reassess which of the intuitive roles we give to "coordinated thinking" are modeled by *thinking with a single vertex* and which by *thinking with a family of vertices connected by certain identity*.

I mean only to acknowledge the option of refining the framework in this direction, not specifically to encourage it. But here is one interesting kind of case that may motivate this refinement. Suppose you are certain that the woman who just left was Carol, but your interlocutor is not, and asks who it was. You reply, "She put on Carol's coat, you figure it out." Though you *do have* coordinated beliefs here about Carol's putting on her own coat and leaving, it's not clear that you *express* a coordinated belief with your reply — though you did believe, and intended to be understood as implying, that "she" and "Carol" corefer. (The literature cited as a "third use of wires" in note 14, above, discuss linguistic mechanisms that, by contrast, are more plausibly dedicated to expressing only coordinated thought.) These cases are complex in part because they deal with you having one "mental file" where your interlocutor has two, and in part because they deal with the connections between coordinated *thinking* and *linguistic* mechanisms. For discussion of the linguistic side of these cases, see Higginbotham 1985, pp. 569ff; Evans 1980, pp. 239–40, 243ff; McKay 1991, pp. 718, 725; Fiengo and May 1994, Chapter. 1; and example 12 in Heim 1998.

I've also remarked on how opposed a neo-Fregean need or needn't be to my framework.

Richard 1990, Fine 2007, and Pinillos 2011 have come up a few times in our discussion, and to those who know those works (and Richard 1983, 1987, and Taschek 1995, 1998), a broad consonance with my framework will also be evident. I will discuss those proposals in detail elsewhere; I haven't addressed them more specifically here because the connections and differences aren't straightforward. I am sympathetic to the semantic ideas they share, but unlike these authors (and Ninan too), I think it's more productive — at least, at this stage of inquiry — to deal with the mental and the linguistic separately. This paper aims only to give a framework for modeling and theorizing about our mental states. What linguistic mechanisms subjects may have to express, or talk about, coordinated thinking, I've deferred for later discussion.

I want to dwell a moment longer on the differences between my framework and a range of nearby positions, that I've called "symbolist" and cited examples of in note 1, above. (See also the symbolist interpretations of mental files and discourse referents in note 31.) These views will invoke particular mental symbols, or words in a language of thought, or private representational vehicles, where I have only posited graph-theoretic structure. The symbolists needn't *resist* much (if anything) of what I've said, since the machinery they invoke *does also have* graph-theoretic structure. They could just be regarded as taking on stronger commitments. To compensate, they can claim greater familiarity, and perhaps more explanatory power.

But are these views really explanatorily better?

I'm not sure exactly what cognitive architecture we commit to when we talk of "mental symbols," but symbolists generally allow that it's possible for subjects to share propositional attitudes while having different, or no, mental symbol systems. And as I said a few times earlier, if we want explanations that are driven by *what we have in common* with other such possible thinkers, we should have models that abstract away from the details of how each subject happens to implement the structures they share. The details of how things are represented in one's own cognitive architecture — not just the neural details, but *whether the implementation has any syntax-like structure at all* — is arguably only important for other, more parochial, explanatory aims.

Consider the following analogy. Suppose I appeal to the algebraic notion of a *ring* for some explanatory purposes, with which you're initially unfamiliar. After learning about rings, you say, "Oh I see, like the integers. Why don't you you just use integers in your model?" I respond: Integers are a more specific kind of structure than a ring. They may be more familiar, and more easily axiomatized (if you're not concerned to emphasize how they are specialized rings and so on). But they are heavier-duty, less minimal structures than rings because of their more specific commitments. If we make explanatory use of integers, we should keep track of when those more specific commitments are appropriate or justified. We have to make sure we don't illegitimately help ourselves to stuff in our model whose presence hasn't yet been fairly motivated.

How does that bear on our present discussion? The symbolists bring more to the explanatory table than just the graph structure of their symbols. But I don't think we're in a position to affirm that these extras are really anything that our explanatory models of the mind, with the aim to capture what we share with other possible thinkers, are entitled to. The "permutation worry" I voiced in Section 9, above, emerges from such doubts. (See also the artifacts cited in note 39.) I expect that making a model of the mind abstract enough to avoid these worries will inevitably have to bring in the kind of graph-theoretical descriptions I've employed, or something that is not substantially any different.

To the extent that the framework presented here really does differ substantially from similar-seeming views, and isn't inter-translatable with them, the take-home story is that my framework aims to be more minimal and abstract, capturing interesting structure the other views posit, while avoiding their commitments about how that structure is implemented. What my framework includes is already enough on its own to model the paradigm Frege phenomena in thought that we've accumulated over the past century and a quarter.

Acknowledgments Thanks to audiences at Barcelona, Paris, Oxford, NYU/La Pietra, Bielefeld, Ohio State, UCLA, Franklin and Marshall, Irvine, and several forums in New York; and especially Chris Barker, Selim Berker, Ned Block, Dylan Bumford, David Chalmers, James Genone, Dan Hoek, Nick Kroll, Andrew Lee, Geoff Lee, Dilip Ninan, Chris Peacocke, Jesse Rappaport, François Recanati, Barry Smith, Aaron Zimmerman, and anonymous reviewers for this journal.

References

Asher, Nicholas. 1986. Belief in Discourse Representation Theory. *Journal of Philosophical Logic* 15: 127–89.

Asher, Nicholas. 1987. A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy* 10: 125–97.

Asher, Nicholas. 1989. Belief, acceptance and belief reports. *Canadian Journal of Philosophy* 19: 321–61.

Austin, David. 1990. *What's the Meaning of "This"?*: Cornell.

Bach, Kent. 1987. *Thought and Reference*: Oxford.

Boër, Steven. 2007. *Thought-Contents: On the Ontology of Belief and the Semantics of Belief Attribution*: Springer.

Chalmers, David. 2003. The content and epistemology of phenomenal belief In eds. Quentin Smith, and A. Jokic, *Consciousness: New Philosophical Essays*, 220–72. Oxford.

Chalmers, David. 2011. Frege's Puzzle and the objects of credence. *Mind* 120: 587–635.

Chalmers, David. 2012. *Constructing the World*: Oxford.

Cresswell, Maxwell, and Arnim von Stechow. 1982. *De re* belief generalized. *Linguistics and Philosophy* 5: 503–35.

Crimmins, Mark, and John Perry. 1989. The prince and the phone booth: reporting puzzling beliefs. *Journal of Philosophy* 86: 685–711.

Crimmins, Mark. 1992. *Talk about Beliefs*: MIT.

Cumming, Sam. 2014. Discourse content In eds. Alexis Burgess, and Brett Sherman, *Metasemantics: New Essays on the Foundations of Meaning*, 214–30. Oxford.

Edelberg, Walter. 1986. A new puzzle about intentional identity. *Journal of Philosophical Logic* 15: 1–25.

Edelberg, Walter. 1992. Intentional identity and the attitudes. *Linguistics and Philosophy* 15: 561–96.

van Eijck, Jan. 2006. Discourse Representation Theory In ed. Keith Brown, *The Encyclopedia of Language and Linguistics, 2nd ed., Vol. 3*, 660–9: Elsevier.

Evans, Gareth. 1977. Pronouns, quantifiers, and relative clauses (I). *Canadian Journal of Philosophy* 7: 76–152. reprinted in Evans 1985, 76–152.

Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry* 11: 337–62. reprinted in Evans 1985, 214–48.

Evans, Gareth. 1982. *Varieties of Reference*: Oxford.

Evans, Gareth. 1985. *Collected Papers*: Clarendon.

Field, Hartry. 1978. Mental representation. *Erkenntnis* 13: 9–61.

Fiengo, Robert, and Robert May. 1994. *Indices and Identity*: MIT.

Fine, Kit. 2007. *Semantic Relationism*: Blackwell.

Fine, Kit. 2010. Comments on Scott Soames 'Coordination Problems'. *Philosophy and Phenomenological Research* 81: 475–84.

Fodor, Jerry. 1975. *The Language of Thought*: Harvard.

Fodor, Jerry. 1978. Propositional attitudes. *Monist* 64: 501–23. reprinted in Fodor 1981, 177–203.

Fodor, Jerry. 1981. *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*: MIT.

Fodor, Jerry. 1990. *A Theory of Content and Other Essays*: MIT.

Fodor, Jerry. 1994. *The Elm and the Expert: Mentalese and its Semantics*: MIT.

Fodor, Jerry. 1998. *Concepts: Where Cognitive Science Went Wrong*: Oxford.

Forbes, Graeme. 1990. The indispensability of *Sinn*. *Philosophical Review* 99: 535–63.

Groenendijk, Jeroen, Martin Stokhof, and Frank Veltman. 1996. Coreference and modality In ed. Shalom Lappin, *The Handbook of Contemporary Semantic Theory*, 179–213: Blackwell.

Halpern, Joseph. 2003. *Reasoning about Uncertainty*: MIT.

Hazan, Allen. 1979. Counterpart-theoretic semantics for modal logic. *Journal of Philosophy* 76: 319–38.

Heck, Richard. 2012. Solving Frege's Puzzle. *Journal of Philosophy* 109: 132–74.

Heim, Irene. 1983. File Change Semantics and the Familiarity Theory of Definiteness In eds. R. Bäuerle, C. Schwarze, and A. von Stechow, *Meaning, Use, and Interpretation of Language*, 164–89: de Gruyter. reprinted in Portner and Partee 2002, 223–48.

Heim, Irene. 1990. E-Type pronouns and donkey anaphora. *Linguistics and Philosophy* 13: 137–77.

Heim, Irene. 1998. Anaphora and semantic interpretation: a reinterpretation of Reinhart's approach In eds. U. Sauerland, and O. Percus, *The Interpretive Tract*, 205–46: MIT Working Papers in Linguistics Vol. 25.

Higginbotham, Jim. 1983. Logical form, binding, and nominals. *Linguistic Inquiry* 14: 395–420.

Higginbotham, Jim. 1985. On semantics. *Linguistic Inquiry* 16: 547–93.

Higginbotham, Jim. 1991. Belief and logical form. *Mind and Language* 6: 344–69.

Johnson, David, and Paul Postal. 1980. *Arc Pair Grammar*: Princeton.

Kamp, Hans. 1981. A theory of truth and semantic representation In eds. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, *Formal Methods in the Study of Language*, 277–322: Mathematische Centrum, University of Amsterdam. reprinted in Portner and Partee 2002, 189–222.

Kamp, Hans. 1984/85. Context, thought and communication. *Proceedings of the Aristotelian Society* 85: 239–61.

Kamp, Hans. 1990. Propositional Attitudes: The Role of Content in Logic, Language, and Mind In eds. C. A. Anderson, and J. Owens, *Prolegomena to a structural account of belief and other attitudes*, 27–90: CSLI Publications.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic*: Springer.

Kaplan, David. 1968. Quantifying in. *Synthese* 19: 178–214.

Kaplan, David. 1986. Opacity In eds. L.E. Hahn, and P.A. Schilpp, *The Philosophy of W.V. Quine*, 229–89: Open Court.

Kaplan, David. 1989. Demonstratives In eds. Joseph Almog, John Perry, and Howard Wettstein, *Themes From Kaplan*, 481–563. Oxford.

King, Jeffrey. 1995. Structured propositions and complex predicates. *Nous* 29: 516–35.

King, Jeffrey. 1996. Structured propositions and sentence structure. *Journal of Philosophical Logic* 25: 495–521.

King, Jeffrey. 2007. *The Nature and Structure of Content*: Oxford.

King, Jeffrey. 2011. Structured propositions. http://plato.stanford.edu/entries/propositions-structured/.

Kripke, Saul. 1980. *Naming and Necessity*: Harvard.

Landman, Fred. 1986. Pegs and alecs In ed. Joseph Halpern, *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the 1986 Conference*, 45–61: Morgan Kaufmann. also in his *Towards a Theory of Information*, Foris 1986, 97–155.

Landman, Fred. 1990. Partial information, modality and intentionality In ed. P. Hanson, *Information, Language and Cognition*, 247–84: University of British Columbia.

Larson, Richard, and Peter Ludlow. 1993. Interpreted Logical Forms. *Synthese* 95: 305–55.

Levine, Joseph. 1988. Demonstrating in Mentalese. *Pacific Philosophical Quarterly* 69: 222–40.

Lewis, David. 1970. General semantics. *Synthese* 22: 18–67. reprinted in Lewis 1983b, 189–229.

Lewis, David. 1979. Attitudes *de dicto* and *de se*. *Philosophical Review* 88: 513–43. reprinted in Lewis 1983b, 133–159.
Lewis, David. 1983a. Individuation by acquaintance and by stipulation. *Philosophical Review* 92: 3–32. reprinted in Lewis 1999, 373–402.
Lewis, David. 1983b. *Philosophical Papers, Vol. 1*: Oxford.
Lewis, David. 1986. *On the Plurality of Worlds*: Oxford.
Lewis, David. 1999. *Papers in Metaphysics and Epistemology*: Cambridge.
Loar, Brian. 1988. Social content and psychological content In eds. Robert H. Grimm, and D. D. Merrill, *Contents of Thought*, 99–110: University of Arizona.
McKay, Thomas. 1991. Representing *de re* beliefs. *Linguistics and Philosophy* 14: 711–39.
Moltmann, Friederike. 2006. Unbound anaphoric pronouns: E-type, dynamic, and structured-propositions approaches. *Synthese* 153: 199–260.
Moore, Robert, and Gary Hendrix. 1982. Computational models of belief and the semantics of belief sentences In eds. S. Peters, and E. Saarinen, *Processes, Beliefs, and Questions*, 107–27: Reidel.
Ninan, Dilip. 2010. *De se* attitudes: ascription and communication. *Philosophy Compass* 5: 551–67.
Ninan, Dilip. 2012. Counterfactual attitudes and multi-centered worlds. *Semantics and Pragmatics* 5(5): 1–57.
Ninan, Dilip. 2013. Self-location and other-location. *Philosophy and Phenomenological Research* 87: 301–31.
Perry, John. 1980. A problem about continued belief. *Pacific Philosophical Quarterly* 61: 317–32. reprinted in Perry 1993, 69–89.
Perry, John. 1993. *The Problem of the Essential Indexical and Other Essays*: Oxford.
Pinillos, Ángel N. 2011. Coreference and meaning. *Philosophical Studies* 154: 301–24.
Portner, Paul, and Barbara Partee. 2002. *Formal Semantics: The Essential Readings*: Blackwell.
Quine, W.V. 1940. *Mathematical Logic*: Harvard.
Recanati, François. 2012. *Mental Files*: Oxford.
Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation*. London: Croom Helm.
Richard, Mark. 1983. Direct reference and ascriptions of belief. *Journal of Philosophical Logic* 12: 425–52. reprinted in Salmon and Soames 1988, 169–96.
Richard, Mark. 1987. Quantification and Leibniz's Law. *Philosophical Review* 96: 555–78.
Richard, Mark. 1990. *Propositional Attitudes: an Essay on Thoughts and How We Ascribe Them*: Cambridge.
Salmon, Nathan. 1986a. Reflexivity. *Notre Dame Journal of Formal Logic* 27: 401–29. reprinted in Salmon and Soames 1988, 240–74.
Salmon, Nathan. 1986b. *Frege's Puzzle*: MIT.
Salmon, Nathan. 1989. Illogical belief. *Philosophical Perspectives* 3: 243–85.
Salmon, Nathan. 1992. Reflections on reflexivity. *Linguistics and Philosophy* 15: 53–63.
Salmon, Nathan. 2006a. A theory of bondage. *Philosophical Review* 115: 415–48.
Salmon, Nathan. 2006b. Terms in bondage. *Philosophical Issues* 16: 263–74.
Salmon, Nathan. 2006c. The resilience of illogical belief. *Nous* 40: 369–75.
Salmon, Nathan. 2010. Lambda in sentences with designators: an ode to complex predication. *Journal of Philosophy* 107: 445–68.
Salmon, Nathan. 2012. Recurrence. *Philosophical Studies* 159: 407–41.
Salmon, Nathan, and Scott Soames. 1988. *Propositions and Attitudes*: Oxford.
Schiffer, Stephen. 1981. Truth and the theory of content In eds. H. Parret, and J. Bouveresse, *Meaning and Understanding*, 204–22: de Gruyter.
Schilder, Paul. 1935. *The Image and Appearance of the Human Body*: Kegan, Paul, Trench, Trubner & Co.
Soames, Scott. 1985. Lost innocence. *Linguistics and Philosophy* 8: 59–71.
Soames, Scott. 1987a. Direct reference, propositional attitudes, and semantic content. *Philosophical Topics* 15: 47–87. reprinted in Salmon and Soames 1988, 197–239.
Soames, Scott. 1987b. Substitutivity In ed. Judith Jarvis Thomson, *On Being and Saying: Essays for Richard Cartwright*, 99–132: MIT.
Soames, Scott. 1989/90. Pronouns and propositional attitudes. *Proceedings of the Aristotelian Society* 90: 191–212.
Soames, Scott. 1994. Attitudes and anaphora. *Philosophical Perspectives* 8: 251–72.
Soames, Scott. 2010. Coordination problems. *Philosophy and Phenomenological Research* 81: 464–74.
Sosa, David. 2010. The Fine line. *Analysis* 70: 347–58.

Spohn, Wolfgang. 1997. On the objects of belief In *Proceedings of the 2nd Conference 'Perspectives on Analytical Philosophy', Vol. 1*, 291–321: de Gruyter.
Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*: Oxford.
Stalnaker, Robert. 2012. *Mere Possibilities: Metaphysical Foundations of Modal Semantics*: Princeton.
Taschek, William. 1995. Belief, substitution, and logical structure. *Nous* 29: 71–95.
Taschek, William. 1998. On ascribing beliefs: content in context. *Journal of Philosophy* 95: 323–53.
Torre, Stephan. 2010. Centered assertion. *Philosophical Studies* 150: 97–114.
Wiggins, David. 1976a. Identity, necessity, and physicalism In ed. S. Körner, *Philosophy of Logic*, 96–132 and 159–82: University of California Press.
Wiggins, David. 1976b. Frege's Problem of the Morning Star and the Evening Star In ed. M. Schirn, *Studies on Frege II, Logic and the Philosophy of Language*, 221–55: Bad Canstatt, Stuttgart.