Granularity problems*

Jens Christian Bjerring & Wolfgang Schwarz

Penultimate draft, forthcoming in *The Philosophical Quarterly*

Abstract

Possible-worlds accounts of mental or linguistic content are often criticized for being too coarse-grained. To make room for more fine-grained distinctions among contents, several authors have recently proposed extending the space of possible worlds by "impossible worlds". We argue that this strategy comes with serious costs: we would effectively have to abandon most of the features that make the possible-worlds framework attractive. More generally, we argue that while there are intuitive and theoretical considerations against overly coarse-grained notions of content, the same kinds of considerations also prohibit an overly fine-grained individuation of content. An adequate notion of content, it seems, should have intermediate granularity. However, it is hard to construe a notion of content that meets these demands. Any notion of content, we suggest, must be either implausibly coarse-grained or implausibly fine-grained (or both).

Keywords Possible-worlds semantics · impossible worlds · hyperintensionality · content · meaning · proposition

*Thanks to two anonymous referees for helpful comments.

1 Introduction

If mental or linguistic content is modelled in terms of possible worlds it becomes impossible to distinguish between necessarily equivalent contents. Since '2+2=4' and 'there are infinitely many primes', for example, are both true at all possible worlds, they come out as having the same content. This is widely regarded as an embarrassment. After all, the two sentences seem to say very different things; they are not cognitively equivalent; they have different communicative effects; they are not interchangeable in attitude reports, subjunctive conditionals, and other embedded contexts.

An attractive strategy to overcome these drawbacks is to extend logical space by "impossible possible worlds", making room for more fine-grained, *hyperintensional* possible-worlds propositions. If there are worlds where 2+2 does not equal 4 and others where there are only finitely many primes, the truth-values of '2+2=4' and 'there are infinitely many primes' no longer coincide throughout the extended space of worlds. The two sentences then express different possible-worlds propositions.

The proposal is not just a technical fix. It also captures the intuition that for ordinary mortals like us, the space of possibilities is larger than classical logical space. We do not know all logical consequences of what we know. Perhaps Goldbach's Conjecture is entailed by the Peano axioms, but we still do not know it. Both the conjecture and its negation are live possibilities for us and both of them can be represented in our language and thought. Yet, in classical logical space, all possible worlds make Goldbach's Conjecture true, if it is true, or they all make it false, if it is false. By including in logical space impossible worlds where the conjecture is true and others where it is false, we seem to account for possibilities that, for us, remain live candidates for actuality.[1]

On closer inspection, it turns out that tracking hyperintensional distinctions with impossible worlds is not as straightforward as it may at first appear. In sections 2 and 3 we argue that we cannot extend the classical possible-worlds framework without giving up some of the core features of that framework. More significantly, in sections 4 and 5 we argue that the theoretical and pre-theoretical roles associated with mental and linguistic content prohibit not only an overly coarse-grained individuation of content, but also an overly fine-grained individuation which, for instance, would see a difference in content for any difference in

[1] Nolan (2014) argues that impossible worlds may have further uses to illuminate metaphysical features of the world such as essence, grounding, and properties.

morphology. An adequate notion of content, it seems, should have intermediate granularity. Yet, given plausible criteria for sameness of content, we will argue that such a notion of content does not exist. Any way of assigning content to linguistic or mental items, we claim, must either be implausibly coarse-grained or implausibly fine-grained, or both.

2 Impossible worlds: semantics

Return to the idea that fine-grained differences in content might be captured by extending the traditional space of possible worlds with impossible worlds. One challenge for this strategy is to construct the relevant space of worlds. Here we will set aside this issue, although we consider it to be an open problem.[2] For concreteness, we will occasionally refer to a toy construction on which possible and impossible worlds are identified with sets of English sentences. A sentence S then counts as *true at* a world w—equivalently, w *verifies* S—iff S is a member of w. This account has obvious difficulties handling ambiguity, vagueness and context-dependence (among other things), but it can serve as a simple illustration for cases where these phenomena can be ignored.[3]

It appears to be a common sentiment among friends of impossible worlds that an extension of logical space with impossible worlds constitutes a fairly conservative or moderate extension of the classical framework. According to Mark Jago, for instance, impossible worlds allow us to incorporate hyperintensionality into the possible-worlds framework 'while preserving its best features' (Jago 2014a, 14). We disagree. We will argue that once impossible worlds are included in logical space, we lose the best features of the possible-worlds framework. In particular, we lose much of the appeal of the framework in natural

[2] See (Jago 2014a) for a detailed discussion. Jago suggests that possible and impossible worlds can be identified with arbitrary sets of sentences in a restricted Lagadonian language $\mathfrak{L}$; an English sentence S is true at a world w iff w contains an $\mathfrak{L}$-translation of S. This construction can run into problems in cases where different English sentences translate into the same $\mathfrak{L}$-sentence. Arguably, 'woodchucks are whistle-pigs' and 'woodchucks are woodchucks' have the same Lagadonian translation. But, as Jago himself admits, there are contexts in which we want to treat these sentences as having different contents.

[3] In the literature on impossible worlds, there is a standard distinction between "American" and "Australian" type impossible worlds (cf. (Berto 2013)). Roughly speaking, truth at an Australian world is closed under some non-classical consequence relation, while American worlds are not subject to any such restriction. For our purposes, only American constructions will be relevant. To model the kinds of possibilities that seem to be live possibilities for ordinary people, we need worlds where A is true but B is false, even when A logically entails B in some non-classical logic. (Unless the non-classical entailment relation is so weak to blur the distinction between Australian and American worlds.)

language semantics, and we have to give up the analysis of belief and information in terms of exclusion of possibilities.[4]

Let's start with semantics—turning to belief and information in section 3. A central task of natural language semantics is to find general, recursive rules that plug into natural language syntax to determine the content of complex expressions based on the contents of simple expressions. In the possible-worlds framework, contents are traditionally identified with functions from worlds to extensions. As a result, the content of a sentence can be specified by stating at which worlds the sentence is true. Classical examples from possible-worlds semantics include the following rules for complex sentences of various types:

(1) A sentence of the form 'A and B' is true at a world w iff A and B are both true at w.

(2) A sentence of the form 'It is not the case that A' is true at a world w iff A is not true at w.

(3) A sentence of the form 'It is necessary that A' is true at a world w iff A is true at all worlds.

(4) A sentence of the form 'It ought to be the case that A' is true at a world w iff A is true at all deontically ideal alternatives to w.

(5) A sentence of the form 'If A were the case then B would be the case' is true at a world w iff B is true at the closest A-world(s) to w (relative to a certain closeness order).

We do not want to defend these particular rules, but we think it is an attractive feature of the possible-worlds framework—and a major reason for its popularity in semantics—that it allows expressing simple and perspicuous rules along those lines. Yet once we extend logical space to account for hyperintensional distinctions, all such rules must be given up.

To illustrate, the above rules for 'and' and 'not' imply that 'it rains' and 'it is not the case that [it is not the case that it rains and it is not the case that [it rains and it is not the case that it rains]]' are true at the same worlds and thus have the same content. This is just what hyperintensional accounts want to avoid. Similarly, on the hyperintensional account, we want to distinguish the content of 'It is necessary that 2+2=4' from that of 'It is necessary that there

[4] We have no objections to the "Australian" use of impossible worlds in the model theory of non-classical logic (see e.g. (Restall 1997)).

are infinitely many primes'. So we cannot say that 'It is necessary that A' is true at a world w iff A is true at all metaphysically possible worlds. Nor, of course, can we say that 'It is necessary that A' is true at w iff A is true at all worlds, whether possible or impossible, since that would render 'It is necessary that 2+2=4' false at the actual world.

More generally, popular rules of natural language semantics reduce the possible assignments of truth-values to sentences. They entail that if 'A and B' is true, then A cannot be false; that if 'someone fears everyone' is true, then 'no one fears themselves' is false; that if the Peano axioms are true, then Fermat's Last Theorem is true. But in the extended space of worlds, we do not want these entailments. We want to allow for worlds where things are true although their consequences are false. In the extended space of worlds, the set of worlds associated with 'It is not the case that A' or 'It is necessary that A' is therefore no longer determined in a systematic and perspicuous way by the set of worlds associated with the embedded sentence A. To be sure, the classical rules may still hold within the restricted space of genuinely possible worlds. But this is little consolation if we are seeking rules that determine the content of complex expressions, given that those contents extend beyond the space of genuinely possible worlds.

Friends of impossible worlds are well aware of the present point. They agree that if we want capture fine-grained differences in content using worlds, then classical rules of possible-worlds semantics must go (as must non-classical alternatives such as the Routley-star semantics for negation). We want to stress that this is a cost. Not only do we have to abandon 70 years of progress in possible-worlds semantics. Worse, hopes of devising new semantic rules that match the power and transparency of the old rules look dim. Most advocates of impossible worlds do not even try to spell out informative rules. David Ripley (2012, 110ff.), for example, suggests that 'and' expresses some function from pairs of sets of worlds to sets of worlds, but he does not say what the function is—and it would be hard to do so. Our toy construction of worlds as sets of sentences allows us to be more specific, but again at a serious cost. In our construction, the meaning of 'and' maps any pair of sets $\{w : A \in w\}$ and $\{w : B \in w\}$ to the corresponding set $\{w : \text{`}A \text{ and } B\text{'} \in w\}$. This operation not only looks nothing like set intersection—the classical interpretation of 'and'. There is also something uninformative and trivial about the new rule: intuitively, it should make a big difference whether 'and' expresses conjunction or disjunction, but the rule just stated is correct either way!

3 Impossible worlds: pragmatics

So far we have argued that extending the possible-worlds framework in order to capture hyperintensional distinctions reduces the power and appeal of the framework for the semantics of natural language. One might respond that the main appeal of the classical possible-worlds framework lies not so much in the resources it provides in formal semantics, but rather in its promise of providing a unified, systematic account covering mental, informational, and linguistic content, and various interactions between these notions.

One starting point here is the idea that to acquire information is to exclude possibilities. When we learn that Bob is in Rome, we can exclude the possibility that he is in Paris. Possibilities—whatever they are—can be ordered by specificity: that Bob is in Rome is a more specific possibility than that he is in Italy; that he is in Rome on a business trip is more specific than that he is in Rome. Under plausible (although non-trivial) assumptions about the specificity ordering, every possibility uniquely corresponds to a set of maximally specific possibilities. (The required assumptions are the conditions on a complete, atomic Boolean lattice.[5]) This is how possible worlds enter the picture: a possible world is a maximally specific possibility, a complete way things might be. We can therefore identify the possibility that Bob is in Rome with a set of possible worlds, and model the information that we receive when we learn that he is in Rome as the set of worlds "at which" Bob is in Rome—a set that excludes (in the set-theoretic sense) all possibilities at which Bob is somewhere else. Similarly, the totality of an agent's knowledge can be modelled as a set of possible worlds, comprising all possibilities that might, for all the agent knows, be actual (see (Hintikka 1962)).

In the same spirit, Robert Stalnaker (1970) proposed a possible-worlds model for the dynamics of assertion. Before Bob told Alice that he is in Rome, the contextually open possibilities did not settle Bob's location; perhaps they included worlds where Bob was in Paris and others where he was in Rome. Bob's utterance of 'I am in Rome' then had the effect that all worlds where he is not in Rome got removed from the set of contextually open possibilities. Had Bob uttered a different sentence—say, 'I am in Paris'—the set of contextually open possibilities would have changed in a different, but equally predictable manner.

[5]The isomorphism between possibilities and sets of maximally specific possibilities is then an instance of Stone's representation theorem—bracketing a technicality concerning non-principal ultrafilters that arises if the space of possibilities is infinite.

The sentences thus have different "context change potential".

These ideas have been successfully extended and refined to analyze a large variety of phenomena. However, as they stand, they are insensitive to hyperintensional distinctions. The possible-worlds model of information and knowledge cannot account for agents who know simple logical truths without knowing all logical truths. Stalnaker's model of assertion seemingly cannot explain the different communicative effects of sentences that are true at the very same worlds. It is here that impossible worlds promise relief.

Yet, as soon as we try to adapt the above picture to an impossible-worlds framework, we run into severe difficulties. Bjerring (2013) points out the following. Assume worlds—whether possible or impossible—are *complete* in the sense that for every sentence they verify either it or its negation. Now consider a world w that verifies some complex contradiction C of, say, classical propositional logic. Since C is a contradiction, there is a proof of $\neg C$. That is, there is a sequence of sentences $S_1, \ldots, S_n$, ending in $S_n = \neg C$, each member of which is either a simple tautology (a propositional "axiom" such as $A \to A$) or derivable from one or two earlier elements in the sequence by a simple logical rule like modus ponens. Given that worlds are complete, w contains either S_i or $\neg S_i$ for each element in the sequence $S_1, \ldots, S_n$. So there are exactly three possibilities for w: either (i) w verifies the negation of some simple tautology, or (ii) w verifies the premises of a simple logical rule as well as the negated conclusion, or (iii) w verifies both C and $\neg C$. In each case, w is a trivially inconsistent world by the standards of classical propositional logic.

So it looks like we cannot allow for situations in which all trivially inconsistent possibilities have been ruled out while some non-trivially inconsistent possibilities remain open. Nor can we model the belief states of logically non-omniscient, yet moderately competent agents who can rule out all trivially inconsistent possibilities without ruling out all non-trivially inconsistent possibilities. To avoid logical omniscience, some inconsistent possibilities must remain live possibilities for such agents—but then some trivially inconsistent possibilities must also remain live possibilities.

The problem is related to what Jago (2014b) calls "the problem of rational knowledge" (see also (Jago 2014a, ch.6)). As Jago points out, if knowledge is not closed under logical consequence then it cannot be closed under trivial consequence either, yet it seems wrong to say of moderately rational agents that they do not know trivial consequences of what they know. In response, Jago suggests that it is a matter of vagueness just how far an agent's knowledge extends through trivial consequence, so that it is never determinately true that an agent knows a proposition but fails to know one of its trivial consequences.

Jago's proposal helps with the problem of rational knowledge, but it is not clear how it helps with the above problem for possible-worlds accounts. If an agent can determinately rule out all trivially inconsistent possibilities, none of the remaining (determinate) possibilities should be trivially inconsistent. In particular, then, none of the remaining maximally specific possibilities – none of the remaining possible worlds – should be trivially inconsistent. By the argument of Bjerring (2013), none of those worlds will therefore verify any complex contradiction; so how can we model agents for whom complex contradictions are possible?

Jago suggests to drop the assumption that worlds are complete. Suppose we allow for incomplete worlds that verify neither S nor $\neg S$, for some sentence S. Such a world can then verify all simple tautologies (for example) but neither verify a trivial consequence S nor its negation $\neg S$. If such worlds are elements of an agent's doxastic space, the agent can fail to be logically omniscient but still rule out any trivially inconsistent possibility.

So far, so good. But recall that in the possible-worlds framework, "possible world" is just a colourful label for a maximally specific possibility. How can a possibility that verifies neither S nor $\neg S$ be a maximally specific ways things might be, by the lights of a moderately rational agent? To be sure, the agent might (rationally or irrationally) believe that S and $\neg S$ are both false. In that case, worlds that verify either S or $\neg S$ are plausibly incompatible with the agent's beliefs. Similarly, if the agent merely reserves some credence for the possibility that S and $\neg S$ are both false, then her space of doxastically possible worlds should include incomplete worlds that verify neither S nor $\neg S$. However, failure of logical omniscience can hardly be reduced to skepticism about bivalence. Consider an agent who is certain that either S_i or $\neg S_i$ is true, for all members of the sequence $S_1, \ldots, S_n$. Worlds that verify neither S_i nor $\neg S_i$ should then not count as live possibilities for her: they are not maximally specific ways things might be. Nonetheless, the complex contradiction C may be deemed possible by the agent, and its negation $\neg C$ may provide her with non-trivial information. This time, the worlds she rules out cannot be incomplete worlds, since those were already ruled out from the start. We are left with the original problem.

The upshot is that we cannot use impossible worlds to characterize the knowledge or belief of logically competent but non-omniscient agents—not if we want to understand these worlds as maximally specific ways things might be. Either the worlds are blatantly inconsistent, in which case they do not represent genuine possibilities (by the lights of the agent), or they are incomplete, in which case they do not represent maximally specific ways things might be, assuming the agent accepts the relevant instances of bivalence.

There is a more general point here. In the classical possible-worlds framework, the identification of possibilities (or propositions) with sets of possible worlds was justified by certain structural assumptions about the space of possibilities, namely that the possibilities ordered by specificity constitute a complete, atomic Boolean lattice. If we move to hyperintensional possibilities, these structural assumptions become highly implausible. In fact, it is not entirely clear what should now count as the specificity order. Is the possibility that Bob is in Rome on a business trip still more specific than that Bob is in Rome? We could say that it is, on the grounds that the former entails the latter. But on this approach, the anti-symmetry condition on the specificity order fails: there will be distinct hyperintensional possibilities that entail one another. Consequently, the possibilities will no longer correspond to sets of maximally specific possibilities (see (Pollard 2008) and (Pollard 2011)). Alternatively, we could try to construe the specificity order in some more fine-grained, hyperintensional manner. For instance, we could say that A counts as more specific than B only if it is *obvious* that B follows from A. But then we lose even more of the structural conditions on a Boolean lattice: the "obvious entailment" relation is not even transitive.

Popular constructions of impossible worlds tend to obscure these facts. In our naive construction of worlds as sets of sentences, there still seems to be a natural correspondence between the hyperintensional possibilities expressed by sentences and sets of worlds, mapping every sentence A to the set of worlds that contain A. However, in sharp contrast to the traditional possible-worlds framework, the entire set-theoretic structure here does no work. The *expressible* propositions—those that correspond to sentences—are always set-theoretically independent: the worlds that verify A never form a subset of the worlds that verify another sentence B. Inexpressible propositions which do stand in non-trivial set-theoretic relations to expressible ones seem to be mere artifacts of the construction. Consider, for instance, the set of worlds that verify '2+2=4' conjoined with a few other worlds that verify '2+2=5'. Is this supposed to be a possibility? Is it a possible object of belief? Do you automatically believe it whenever you believe that 2+2=4?

What is really modelled by the hyperintensional construction is an account on which different contents are simply independent entities, and where learning something amounts to *adding it* to a stock of previously learned contents. In our simple construction, when we model an agent's belief state by a set of worlds, the only aspect of the model with real significance is which sentences are verified by all worlds in the set. These are the sentences the agent believes. When the agent learns another sentence, the set of sentences verified by all worlds grows by one. Superficially, learning A is still modelled as excluding worlds that do not verify A. But these worlds are not maximally specific ways things could be. They are not possibilities at all. The exclusion operation on the space of worlds is just a roundabout way of representing the addition of a new sentence.

All this need not show that the hyperintensional account does not work. It only shows that the account should not be regarded as a moderate extension of the traditional possible-worlds account. It is a completely different approach, disguised as a moderate extension.

Now one might argue that a radically different approach is indeed required to account for the hyperintensionality of mental and linguistic content. But before we jump to a new approach, let us try to get clear about the goal. If possible-worlds propositions are too coarse-grained, what grain size would be adequate?

4 The need for coarse-graining

Our toy construction of worlds as sets of sentences makes possible-world propositions extremely fine-grained. No two sentences ever express the same content, since there are always worlds that contain one but not the other. However, an excessively fine-grained individuation of content is as problematic as an excessively coarse-grained one.

To begin, intuitively we do think that at least some sentences have the same content. '2+2=4' and 'there are infinitely many primes' say quite different things, but that is not true for pairs such as 'I nearly fell' and 'I almost fell', for '$3 < 9$' and '$9 > 3$', or for 'it is raining' in English and 'il pleut' in French. In these pairs, both sentences intuitively express the very same thought and make the same claim about reality.

Second and more seriously, consider Euclid's discovery that there are infinitely many primes. Presumably the content of Euclid's discovery can be expressed not only by 'there are infinitely many primes' but also by trivially equivalent statements such as 'the number of primes is infinite': that the number of primes is infinite was not a further discovery, also made by Euclid. Similarly, when the Babylonians discovered that the morning star is identical to the evening star, the content of their discovery is equally expressed by 'the evening star is the morning star'. In each case, the content of the discovery seems to have intermediate granularity.

The same is true for other attitudes. You cannot notice that I nearly fell without noticing that I almost fell. But it is not like there is a mysterious necessary connection between the two noticings. Rather, there is only one state of noticing that is equally described as 'noticing that I nearly fell' and 'noticing that I almost fell'. So if noticing is a relation between a subject and a content, then the relevant content is not as fine-grained as the words we can use to express it.

Third, overly fine-grained conceptions of content seem to preclude sentences in different languages from having the same content. If no two English sentences agree in content, then it is hard to see how sentences from different languages could achieve that feat—especially given that translations into other languages often involve changes in grammatical structure and sub-sentential meaning, such as the translation from 'it is raining' to the Russian 'Идёт дожд' (literally, 'goes rain'). But the hypothesis that sentences in different languages never agree in content is not only intuitively implausible, it also leads to further problems for attitude reports. If 'S believes that A' attributes a belief whose content is that of the embedded sentence A, then 'Euclid believed that there are infinitely many primes' would attribute to Euclid a belief that can only be expressed or attributed in English—which would raise the question how Euclid could have acquired such an "essentially English" belief, since neither he nor anyone else at his time spoke English.

This last remark leads to a fourth argument against excessively fine-grained conceptions of content: such conceptions make it mysterious how mental and linguistic types get to have their content. Suppose the content of a mental state is determined by causal relations to the environment, behavioural dispositions, inferential links, and further features along these lines. It is then hard to see how there could be a genuine difference in content for any two sentences that might be used to attribute an attitude.[6] And if mental content is relatively coarse-

[6]Stalnaker (1976) even argues that functionalist accounts of mental content only deliver the very coarse-grained contents of traditional possible-worlds accounts.

grained, then many foundational accounts of linguistic content—for example, in the tradition of (Grice 1957)—imply that linguistic content will be equally coarse-grained.

Fifth and finally, intermediately-grained notions of content are central to many projects in philosophy. For example, a key idea in Frege's *Grundlagen* (1884) is that quantified mathematical statements sometimes agree in content with non-quantified statements, as in the case of (*) and (**):

(*) A and B are parallel.

(**) A and B have the same direction.

Frege's claim would be trivially false if syntactically different sentences could never agree in content.

One problem that emerges from all these considerations is that we will probably not find a single level of granularity that works for every purpose, in every context. The kind of content established by a formal proof in, say, intuitionistic logic, is a lot more fine-grained than the content of an astronomical discovery or the content represented by a map. When we talk about beliefs, we often want to treat the propositions that there are woodchucks and that there are whistle-pigs as identical, especially when the subject of the belief does not speak English; but not always (see (Ripley 2012)). This suggests that there is no way of assigning contents to sentences that will get all cases right.[7]

Frege put forward a criterion for identity of content that—albeit inadvertently—takes into account some such flexibility and relativity. In essence, Frege's proposal is that two sentences have the same content if and only if one could not regard one of them as true and the other as false. Sentences that satisfy this condition Frege called *equipollent* (see e.g. (Frege 1891, 14), (Frege 1892, 47), and (Frege 1983, 152f.)).[8] By Frege's criterion, '2+2=4' and 'there are infinitely many primes' plausibly come out as having different contents, while '$3 < 9$' and '$9 > 3$', or (*) and (**) have the same content.

In general, however, whether two sentences count as equipollent will depend on what kinds of agents we consider when we ask whether it is possible to hold that the sentences have different truth-values. The individuation of content becomes relative to a base level of information and cognitive capacities. Consider

[7] Stalnaker raises versions of the present worry in (Stalnaker 1991) and (Stalnaker 1999b).

[8] We will not enter into the exegetical details about what exactly Frege meant by equipollence and how his proposal squares with other remarks in which he seems to suggest that the sense of a sentence is composed of the senses of its parts; see e.g. (Penco 2003), (Kemmerling 2010), and (Schellenberg 2012) for discussion.

(*) and (**). Most speakers of English would probably dismiss the hypothesis that one of these might be true while the other is false, but it is not hard to imagine speakers for whom this is a live possibility. On a somewhat more advanced level, consider the equations '$x = \ln y$' and '$y = e^x$'. While an algebra student may wonder whether they can differ in truth-value, her teacher may find it utterly obvious that they cannot. In the limit, if we only consider ideal agents with unbounded cognitive capacities, equipollence might reduce to something like necessary or a priori equivalence. At the other extreme, for utterly confused agents—or, less interestingly, for agents who do not speak the relevant language—it reduces to a trivial relation that never holds between different sentences.

We can therefore see Frege's proposal as offering a whole range of criteria whose end-points are the extremely coarse-grained individuation of classical intensional semantics and the extremely fine-grained individuation of our toy construction. Since it is doubtful that a single grain size is adequate in all contexts and for all purposes, some such pluralism about content might be just what we need.

Now Frege's criterion only tells us when two sentences have the same content. It does not tell us what these contents are. Unfortunately, we will see that it is impossible to construe a notion of content that satisfies Frege's criterion for intermediary points along the scale, where we consider the judgements of moderately competent, but logically non-omniscient agents.

5 The intransitivity of sameness of content

Consider another lengthy sequence $S_1, \ldots, S_n$ of sentences. This time assume that each S_{i+1} is trivially equivalent to its predecessor S_i—meaning that each trivially follows from the other—although S_n is not trivially equivalent to S_1. For example, $S_1, \ldots, S_n$ might be a sequence of algebraic equations, where each step is a trivial transformation of the previous equation, although S_n is a highly non-trivial transformation of S_1. In section 3, we asked how possible-worlds models can take into account the fact that moderately rational agents may know S_1 without knowing S_n. In the present section, we will not assume the possible-worlds model, and our interest is not in modelling knowledge. Rather, we want to ask which of the sentences $S_1, \ldots, S_n$ agree in content.

Can we say, following Frege, that S_i and S_j agree in content iff moderately competent agents can rule out the hypothesis that S_i and S_j differ in truth-value? The condition may well be satisfied for all subsequent equations in our sequence: presented with adjacent sentences, moderately competent agents can immediately see that they are equivalent. On the other hand, they cannot immediately see that first and the last equation are equivalent; it remains a live possibility for them that the first equation is true and the last one false. So S_1 and S_n are not equipollent. By Frege's criterion, it follows that each sentence in the sequence has the same content as its immediate neighbours, although the first and the last sentence do not. Evidently, this is impossible.

Note that the problem here does not rely on any assumptions about the nature of contents. It does not matter whether contents are sets of worlds, structured entities, mental states, or *sui generis* whatnots. The problem is that equipollence is intransitive, while identity is transitive. So it cannot be true, as Frege's criterion requires, that two sentences have identical content just in case they are equipollent. The problem disappears only at the end points of the Fregean spectrum. For ideal agents who instantly and effortlessly recognize every consequence of every sentence, equipollence may well be transitive. The same is true for utterly confused agents for whom no sentence is equipollent to any other.

It would be short-sighted to blame Frege's criterion. The general problem does not turn on Frege's particular individuation of intermediately-grained content. Return to our sequence $S_1, \ldots, S_n$. Any assignment of content must cut the sequence into equivalence classes of sentences with the same content. If we are looking for an intermediately-grained notion of content, we do not want too many cuts in the sequence. In particular, we do not want to say that no two sentences in the sequence have the same content. We also do not want to say that all sentences have the same content. But then it gets hard to justify the cuts. Suppose the first cut is after S_{10}. So S_1 and S_{10} count as having the same content, while S_{10} and S_{11} count as having different contents, despite the fact that S_{10} and S_{11} are more obviously equivalent than S_1 and S_{10}.

Resorting to unsharp cuts would not help. Suppose we accept a fuzzy notion of content on which it may be vague or indeterminate whether two sentences have the same content, perhaps corresponding to the vagueness in Frege's characterization of equipollence. By the considerations of the previous section, *some* sentences should still determinately agree in content. So let $S_1, \ldots, S_n$ be a sequence where neighbouring sentences determinately agree in content, but the endpoints determinately have different contents. Since determinately having the same content is transitive, our problem remains: there is no fuzzy assignment of contents, and no assignment of fuzzy contents, on which neighbouring sentences determinately have the same content and yet distant sentences do not.

The problem also arises when we consider sentences $S_1, \ldots, S_n$ across different languages, where each pair of adjacent sentences seems to agree in meaning, although S_1 and S_n do not. For a simple example, consider a context in which we want to distinguish between the proposition that there are woodchucks and the proposition that there are whistle-pigs. In German there is only one word for woodchucks: 'Waldmurmeltier'. So both 'there are woodchucks' and 'there are whistle-pigs' translate into 'es gibt Waldmurmeltiere'. Now which of these three sentences have the same content?

For another variation, consider a sequence of belief reports $R_1, \ldots, R_n$—in different languages, perhaps—where in each adjacent pair the complement sentences that specify the believed content are trivially equivalent, although the complement of R_1 is not trivially equivalent to that of R_n. Again, we face the same uncomfortable choice between saying that practically any change in complement sentence attributes a different belief, or making isolated cuts where a small change in complement sentence amounts to a different belief, even though other, intuitively larger changes do not.

Of course, there are systematic ways of placing the cuts. The classical neo-Russellian account, for example, sees a difference in content whenever there is either a difference in syntactic structure or a difference in reference. Consequently, 'Hesperus is Hesperus' and 'Hesperus is Phosphorus' are assigned the same content, while '$3 < 9$' and '$9 > 3$', or 'it is raining' and 'Идёт дожд' are assigned different contents. But this way of individuating content does not fit any of the phenomena surveyed in the previous section that seemed to call for an intermediately-grained notion of content. By ordinary standards, for example, discovering that Hesperus is Phosphorus is not at all the same thing as discovering that Hesperus is Hesperus, while it would be perfectly fine to describe an utterance of 'Идёт дожд' as an assertion that it is raining.

The same is true for David Chalmers' recent proposal in (Chalmers 2011) and (Chalmers 2012, 248ff.) on which two sentences have different contents iff they either differ in syntactic structure or some of their constituents differ in (primary) intension or extension. 'Hesperus is Phosphorus' and 'Hesperus is Hesperus' now plausibly come out as having different contents. However, '$3 < 9$' and '$9 > 3$', 'it is raining' and 'Идёт дожд', or Frege's 'A and B are parallel' and 'A and B have the same direction' also have different contents.

(*Pace* Chalmers, his notion of content is therefore not very Fregean.) Looking back at the phenomena from the previous section, Chalmers' notion of content almost always cuts too finely. Sometimes it cuts too coarsely. For example, if we define 'ν' to denote the smallest positive number x for which $\cos(x/2) = 0$, then '$\nu = \pi$' and '$\pi = \pi$' have the same Chalmersian content, although the former seems informative but the latter trivial.[9]

The argument we have outlined is completely general, so there is no need to survey other proposals. By the intuitive and theoretical considerations we have reviewed, there are sequences of sentences in which adjacent sentences should (determinately) have the same content while sentences that are sufficiently far apart should (determinately) have different contents. No assignment of contents to sentences can satisfy this requirement.

6 Conclusions

We have tried to establish two main claims. First, the classical possible-worlds framework is essentially coarse-grained: while fine-grained contents can be formally identified with sets of possible or impossible worlds, the resulting propositions can no longer play the role of classical possible-worlds propositions in semantics and models of knowledge, information, and communication.

Second, although a variety of phenomena seem to call for a notion of content that is more fine-grained than sets of possible worlds and more coarse-grained than linguistic morphology, there are serious obstacles to construing an adequate notion of intermediately-grained content. For one, different phenomena and different contexts seem to call for different levels of granularity. Worse, there is no possible assignment of contents to sentences that matches intuitively plausible criteria for sameness of content such as Frege's equipollence principle. The reason is simple: content identity is transitive but the criteria make sameness of content intransitive. Given these problems, it is no surprise that popular accounts of linguistic and mental content tend to be either implausibly

[9] Chalmers' individuation of content resembles Carnap's individuation in terms of intensional isomorphism. The present objection to Chalmers is raised in (Church 1954) as an objection to Carnap. Chalmers mentions the problem in (Chalmers 2012, 249) and replies that 'it is at least arguable that ['ν'] should be understood to have complex structured content', which would distinguish it from 'π'. But then the relevant structure cannot be tied to syntax or logical form, which is what Chalmers' official proposal assumes. If even syntactically simple terms can have structure, one would like to know a lot more about how that structure is determined. Does 'π' have structured content? Which of the many equivalent definitions of π is reflected in its structure?

coarse-grained or implausibly fine-grained (or both).

This leaves two options for semantic theorizing. One is to drop the assumption that semantic facts can be captured by assigning to linguistic or mental items some kind of extra-linguistic content. If instead we confine ourselves to studying relations within the domain of the linguistic or the mental—relations of same-saying, synonymy or equipollence, for example—then it does not matter whether these relations are transitive or context-dependent. Alternatively, we can continue assigning content to linguistic and mental items, accept that our assignment is implausibly coarse-grained or implausibly fine-grained (or both), and try to explain away the phenomena that seem to call for an intermediately-grained notion of content.

An example of the explaining-away strategy is Stalnaker's appeal to "metalinguistic diagonalization" in his account of assertion (e.g. (Stalnaker 2004)). According to Stalnaker, what is asserted by an utterance of 'Hesperus is Phosphorus' is the set of all worlds. However, when we encounter such an utterance in otherwise ordinary circumstances, general pragmatic rules make us realize that what the speaker tries to communicate is the contingent proposition that whichever heavenly body is picked out by 'Hesperus' is also picked out by 'Phosphorus'. Salmon (1986) likewise offers a pragmatic explanation of why one apparently cannot always replace 'Hesperus is Phosphorus' and 'Hesperus is Hesperus' in attitude reports, even though (on Salmon's account) the two sentences express the same proposition.

We are skeptical that purely pragmatic accounts will suffice to explain away all the phenomena that seem to call for intermediately-grained content. A more promising approach, we believe, is to drop the assumption that attitude reports and speech act reports simply state a relation between the subject and the content expressed by the embedded sentence. Without that assumption, the hypothesis that 'Hesperus is Phosphorus' and 'Hesperus is Hesperus'—or '2+2=4' and 'there are infinitely many primes'—express the same content no longer makes the false prediction that the two sentences are interchangeable in attitude reports or speech act reports: it no longer follows that anyone who believes (or asserts) that 2+2=4 thereby also believes (or asserts) that there are infinitely many primes.

On any approach, we should stop faulting extant accounts of content for being implausibly coarse-grained or fine-grained. No conception of content—no matter how coarse-grained or fine-grained—can fit the identity conditions apparently imposed by ordinary judgements about meanings, attitudes, and speech acts.

Aarhus University, Denmark
The University of Edinburgh, Scotland

References

Francesco Berto [2013]: "Impossible Worlds". In Edward N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy,* Winter 2013 edition

Jens Christian Bjerring [2013]: "Impossible worlds and logical omniscience: an impossibility result". *Synthese*, 190(13): 2505–2524

David Chalmers [2011]: "Propositions and Attitude Ascriptions: A Fregean Account". *Nous*, 45: 595–639

— [2012]: *Constructing the World*. Oxford: Oxford University Press

Alonzo Church [1954]: "Intensional Isomorphism and Identity of Belief". *Philosophical Studies*, 5(5): 65–73

Gottlob Frege [1884]: *Die Grundlagen der Arithmetik*. Beslau: M.& H. Marcus

— [1891]: *Funktion und Begriff*. Jena: Hermann Pohle

— [1892]: "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50

— [1983]: *Nachgelassene Schriften*. Hamburg: Meiner, 2nd edition

Paul Grice [1957]: "Meaning". *The Philosophical Review*, 66(3): 377–388

Jaakko Hintikka [1962]: *Knowledge and Belief: An introduction to the logic of the two notions*. Ithaca: Cornell University Press

Mark Jago [2014a]: *The Impossible*. Oxford: Oxford University Press

— [2014b]: "The Problem of Rational Knowledge". *Erkenntnis*, 79(S6): 1151–1168

Andreas Kemmerling [2010]: "Thoughts without parts: Frege's doctrine". *Grazer Philosophische Studien*, 82: 165–188

Daniel Nolan [2014]: "Hyperintensional Metaphysics". *Philosophical Studies*, 171(1): 149–160

Carlo Penco [2003]: "Frege: two theses, two senses". *History and Philosophy of logic*, 24(2): 87–109

Carl Pollard [2008]: "Hyperintensions". *Journal of Logic and Computation*, 18(2): 257–282

— [2011]: "Are (Linguists') Propositions (Topos) Propositions?" In *Logical Aspects of Computational Linguistics,* Springer, 205–218

Greg Restall [1997]: "Ways Things Can't Be". *Notre Dame Journal of Formal Logic*, 38: 583–597

David Ripley [2012]: "Structures and circumstances: Two ways to fine-grain propositions". *Synthese*, 189(1): 97–118

Nathan Salmon [1986]: *Frege's Puzzle*. Cambridge (Mass.): MIT Press

Susanna Schellenberg [2012]: "Sameness of Fregean sense". *Synthese*, 189(1): 163–175

Robert Stalnaker [1970]: "Pragmatics". *Synthese*, 22. In Stalnaker 1999a

— [1976]: "Propositions". In A.F. MacKay and D.D. Merrill (Eds.) *Issues in the Philosophy of Language,* New Haven: Yale University Press, 79–91

— [1991]: "The Problem of Logical Omniscience I". *Synthese*, 89

— [1999a]: *Context and Content*. Oxford: Oxford University Press

— [1999b]: "The Problem of Logical Omniscience II". In Stalnaker 1999a, 255–273

— [2004]: "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics". *Philosophical Studies*, 118: 299–322