Conceptual Role Semantics in a Spatial Probabilistic Interpretation

Class Paper for Logic and Probability
Prof. Dr. Hannes Leitgeb
SS2017, LMU Munich

Conrad Friedrich
conradfriedrich@posteo.net

Word count: 5613

September 22, 2017

Contents

1 Introduction

With the goal to give a convincing explication of meaning of linguistic expressions by way of mental representation, spatial representations have been used by quite a few authors. In general, the idea of a spatial representation of mental representation, especially regarding the semantics of (natural) languages, is not a new one. In particular Gärdenfors and Churchland use this metaphor in very distinct ways. This paper starts by skimming[1] over Gärdenfors' conceptual spaces and Churchland's semantic state spaces. Then Field's account of conceptual role quantified by conditional probabilities is summarized. Realizing that Field's notion lends itself to a spatial interpretation, an original approach to tying semantics to spaces based on conditional probabilities is presented. We precisify the idea, develop a formal account of conceptual similarity and discuss the consequences of the account when evaluating context-dependent meaning as well as when trying to match inter-speaker meaning. Concluding, we look at salient objections and point to directions of future research.

1.1 Gärdenfors, Churchland and Field on Spatial Semantics

Peter Gärdenfors is motivated by results from cognitive science and aims to give a substantial and comprehensive account of mental representation and meaning.[2] His account uses conceptual spaces to create somewhat of a middle ground between reductivist connectionist and symbolic system approaches to mental representations. The spaces are intended to provide a general framework. The motivation is, as it strikes me, first and foremost to make sense of descriptive elements. Objects of cognitive experiences get located in spaces determined by qualities belonging to the phenomenal or scientific realm, such as temperature, weight, brightness, pitch and so on. A single object of experience can be located in many such spaces, each making up a domain. There is a specific space for color experience, for example, hue, chromaticness and brightness make up its dimensions. Gärdenfors' dimensions are rooted in phenomenal experiences and mostly are motivated by empirical, psychological research. Concepts can span many different domains, for example is a concept 'bird' represented in many domains, as a bird has weight, colors, makes a sound etc.

Notably, his conceptual spaces are such that concepts correspond to *areas* or volumes,

1. The literature is a bit too vast and deep-reaching for a term paper to firmly put the account presented here in a definite place in the wider landscape, such that the introduction may only serve as a very vague positioning. It would be quite interesting to properly analyze the various theories of mental representation and semantics to see how this account may fit into the bigger picture.

2. Most notably cf. Gärdenfors (2004) and Gärdenfors (2014).

whereas objects or phenomenal stimuli correspond to points. To contrast: The account proposed in this paper represents the conceptual role of a linguistic expression as a point in a spatial representation.

Paul M. Churchland's spatial representation are based on what may be called neural activation patterns. States of a neural structure, such as actual brain modules or artificial neural networks, can be described by analyzing the state of each neuron. A vector of these neuronal states live in a state space, which is the set of all possible state combinations of neuronal states for a given structure, as Tiffany (1999) summarizes. Churchland's claim then is that these states may actually be a vehicle for content, that is, carry mental representations. They also may be interpreted as carrying *semantic* representations, as Fodor and Lepore (1999) suggest, which they immediately criticise. The debate is long and goes through several iterations, such that an attempt at a comprehensive overview here is futile. Instead, let me give a quote of Churchland's that echos Fodor and Lepore's criticism:

> What was presented as new and interesting is therefore something quite old and boring, runs the objection. Worse, the emerging account leaves unaddressed the question of how each axis of the activation space, or each neuron of the representing population, gets its semantic or representational content.[3]

Interestingly, it is a very similar problem that this paper's account has to face when trying to make sense of inter-speaker similarity of meaning.

Geometric approaches have been extraordinarily popular in Computational Linguistics and Natural Language Processing in recent times[4]. The approach here is much more empirical, as the linguists work with actual *data*. The enormous success of these ventures sheds some positive light on the viability of the whole idea of geometric explications of meaning and, in particular, conceptual role.

Common in Computational Linguistics is the statistical analysis of huge amounts of data in text form, assigning quantifiable features to linguistic expressions in this data. The goal here is to find some statistical patterns in the texts that indicate relevant properties of the linguistic expressions, for example recognition of the syntactic role of the expression, named entity recognition, or whether the author of a text talks about her subject favourably or dismissively (so called sentiment analysis). All these tasks have been met with quite the success, suggesting that there is a tangible connection of structural information represented geometrically and what might be generously called

3. Churchland (1998, 7)

4. cf, e.g. Erk and Padó (2008), Sahlgren (2006).

the *content* of the data. The practical usefulness of the geometric interpretation is a hint to its potential theoretical fruitfulness.

Hartry Field[5] proposes that the meaning of a sentence is made up of *referential meaning*, i.e. its truth conditions given by a standard correspondence theory of truth, and *conceptual role*, i.e. its evidential relations to other sentences. The referential meaning serves to "to make questions of intersubjective synonymy *in large part* objective" (Field 1977, 399).

To rationally reconstruct the conceptual role, Field supposes a propositional language and a speaker's subjective probability function. The conceptual role of a sentence is then given by its evidential relations to other sentences, such that two sentences have the *same* conceptual role for this speaker if they are judged equally probable given any sentence of the language.

To tie these accounts together, one may note that all of them say something about *meaning* in spatial reconstructions. Field seems to adhere to a very traditional extensional explication of meaning, in the sense that the truth conditions for sentences at the actual world seem to play a crucial role in what he calls referential meaning. That is only part of his account, though, and the conceptual role part hints at a more cognitivist aspect of meaning. Conceptual role, as described here, supervenes after all on subjective probabilities, and those are idealisations of mental states. Gärdenfors is a proponent of cognitivist semantics, where linguistic expressions refer to something in the mind of the speaker (Gärdenfors 2004, 154). The association with an external world is then just a matter of pragmatics, and a successfully interacting agent has an appropriate conceptual structure. The account used in computational linguistics is mostly functionalist, limiting the meaning of expressions solely to their communicative role.

In any case, there is substantial support from the literature that it's sensible to pursue a spatial account of conceptual role as a way to explicate semantic meaning, which is what the next section starts with.

2 Conceptual Similarity in Probabilistic Space

It's now a small mental leap to the account of conceptual similarity proposed in this paper.

Speaking loosely, two sentences are similar in conceptual role if they agree in subjective probability conditional on most sentences, bar a few outliers, or if they mostly agree conditional on all sentences. To make things precise it would be nice to have a *distance*

5. cf. Field (1977).

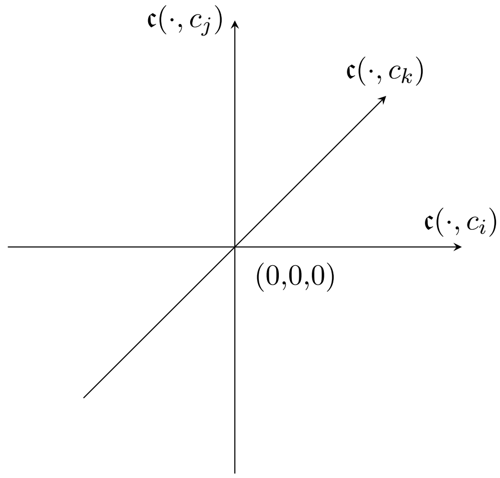

Figure 1: Graphical representation of 3 dimensions of V.

measure on conceptual role of sentences (and, possibly, other linguistic entities) relative to a given subjective probability function, not unlike the accuracy measures employed in justification strategies for Bayesianism.[6]

Presupposing subjective probabilities over a language, Field captures sameness in conceptual role (and *a fortiori* meaning) of two sentences through comparing their probabilities given each sentence of the language. Yet, the result is only a simple binary categorization. Either two sentences have the exact same conceptual role or they don't. While I find the idea of evidential relations determining the conceptual role of a linguistic expression to be very engaging, Field's treatment leaves some things to be desired. There may be expressions with the same conceptual role in a language, that is for sure, but what about all the other expressions? Don't two sentences with *almost* the same conceptual role exhibit an interesting relation to one another? As much has already been deemed plausible by Leitgeb (2008), quoting Goodman (1972). What about whole clusters of expressions? What about the comparison of conceptual roles of sentences between speakers?

I argue that there is rather interesting information of this type contained in the presupposed subjective probabilities. In this paper, I want to present the general idea, make it somewhat precise and sketch directions for future research.

6. cf. Leitgeb and Pettigrew (2010).

2.1 The Central Formalism

Suppose a finite propositional language $\mathcal{L}$ and an agent's subjective probability function $\Pr : \mathcal{L} \rightarrow [0,1]$.[7] A spatial representation of the conceptual role of the sentences of the language can be achieved by assigning *sentences* to dimensions in a space. Let V be an n-dimensional vector space over $\mathbb{R}$, where n is the number of sentences in $\mathcal{L}$. To determine the vector of $p \in \mathcal{L}$ in V, we measure the *evidential relation* of a sentence c_i to p for each dimension, that is, each $c_i \in \mathcal{L}$.

A plausible assumption we adopt from Field is that the evidential relation is captured at least partly by the conditional probability $\Pr(p \mid c_i)$, which consequently takes center stage here[8]. To express the evidential relation in a single number isn't trivial or uncontested (cf. Section 2.4), but let's suppose there is such a measure $\mathfrak{c} : S \times S \rightarrow \mathbb{R}$.

The vector of p in the spatial interpretation $\vec{\mathrm{p}} \in V$ is then fully determined by $\vec{\mathrm{p}} = (\mathfrak{c}(p, c_1), \ldots, \mathfrak{c}(p, c_n))$.

To say something about the relation of the conceptual role of two different sentences $p, q \in \mathcal{L}$, we assume $\vec{\mathrm{p}}$ and $\vec{\mathrm{q}} \in V$ and introduce a second measure $\mathfrak{d} : V \times V \rightarrow \mathbb{R}^+$ for the distance between $\vec{\mathrm{p}}$ and $\vec{\mathrm{q}}$. This can be achieved, among alternatives, by letting $\mathfrak{d}(\vec{\mathrm{p}}, \vec{\mathrm{q}}) = \|\vec{\mathrm{p}} - \vec{\mathrm{q}}\|$ where

$$\|\vec{\mathrm{p}} - \vec{\mathrm{q}}\| = \left(\sum_{k=1}^{n} (p_k - q_k)^2 \right)^{1/2},$$

which is just the euclidean distance.

The *similarity* between two sentences is then just a scoring function that decreases as the distance increases and has a maximum when the distance is minimal, for example: $s_{pq} = e^{-cd_{pq}}$, with a scalar c, and where d_{pq} is the distance of two vectors $\vec{\mathrm{p}}, \vec{\mathrm{q}} \in V$.

In general we have a function $\mathfrak{f} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ that maps distances to similarity measures. We require a monotonically decreasing, continuous function to make sense of the orderings by similarity in the next section.

Summing up, measuring the evidential relation of a sentence to all other sentences of the language with the help of a function $\mathfrak{c}$ generates a representation of that sentence in a vector space. Similarity scores are assigned to pairs of vectors with the composition of distance measure and similarity score: $\mathfrak{f} \circ \mathfrak{d} : V \times V \rightarrow \mathbb{R}^+$.

What is won with this precisification? It creates a precise notion of the spatial inter-

7. For simplicity, $\mathcal{L}$ here just is a set of sentences.

8. Where defined. Field employs Popper functions in his paper, but since this breaks with the orthodoxy and they mostly prove useful when treating infinite sets (Leitgeb 2013, 1352), I assume a classically axiomatised probability function, but not much hangs on this.

pretation of similarity in conceptual role given an agent's subjective probability function. If we additionally buy into Field's premise of meaning being mainly constituted by conceptual role, the spatial interpretation can tell us quite a bit about the meaning of linguistic expressions, as will be seen in the next section.

2.2 Order by Conceptual Similarity

Let's fix a sentence p. The spatial structure of the similarity measure induces relations on $\mathcal{L}$ with interesting properties. To start, we can define

$$q \leqslant_p r :\Leftrightarrow s(p,q) \leqslant s(p,r)$$

for any two sentences q and r. In other words, q stands in the $\leqslant_p$-relation to r if and only if q is less similar in conceptual role to p than r is similar to p. Inducing an ordering on $\mathcal{L}$ that is *total*, *reflexive*, *transitive* but *not* antisymmetric since two sentences q and r may be equidistant to p, written $q \leqslant_p r$ and $r \leqslant_p q$, while $q \neq r$. Less formally, this relation orders all sentences by conceptual similarity to p. Naturally, among the most similar sentence to p is p itself, so it holds that for all sentences $x \in \mathcal{L}$, $x \leqslant_p p$.

Additionally, ſoδ induces equivalence relations on $\mathcal{L}$, so that we can define for sentences q and r

$$q \sim_p r :\Leftrightarrow q \leqslant_p r \wedge r \leqslant_p q,$$

dividing up all sentences into equidistant equivalence classes relative to p.

A neat consequence is another induced ordering: Consider the set of all equivalence classes $\mathcal{L}/\sim_p = \{q/\sim_p \mid q \in \mathcal{L}\}$. We define

$$q/\sim_p \quad <_p \quad r/\sim_p \quad :\Leftrightarrow q \leqslant_p r \wedge r \nleqslant_p q$$

and since this is a relation on equivalence classes, the resulting ordering is strict, that is, not reflexive. It is still transitive, of course.

Now, let $l_1, \ldots, l_m$ list all elements of $\mathcal{L}/\sim_p$ such that $l_i <_p l_{i+1}$. By definition, we have for $a \in l_i, b \in l_{i+1}$ that $a \leqslant_p b$ for all $0 < i < m$. Geometrically, all elements of l_i are equidistant to p and thus lie on the surface of the same hypersphere. This hypersphere contains all elements of l_j with $j < i$, such that we can define sets $S_p^1, \ldots, S_p^m$ with

(i) $S_p^1 = l_1$, with $p \in S_p^1$ and

(ii) [9] $S_p^i \subset S_p^{i+1}$.

9. The inclusion is strict since we already excluded equivalence through the ordering on equivalence

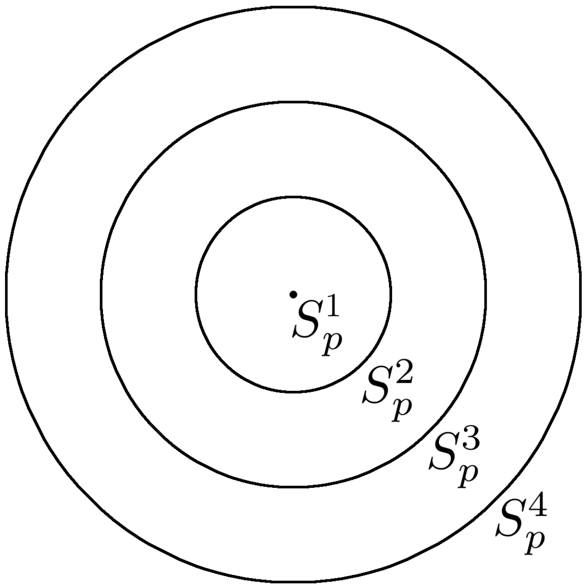

Figure 2: A two-dimensional representation of n-dimensional conceptual similarity spheres for a simple language with 4 equivalence classes

Or in other words, $S_p^{i+1} \setminus S_p^i = l_{i+1}$. See Figure 2 for a graphical representation which shows different levels of similarity to our fixed sentence, p. The most similar in conceptual role is p itself.[10] S_p^2 contains all second-closest sentences and so forth. The *distance between* two sentences in $S_p^{i+1} \setminus S_p^i$ can be greater or less than to p, but is limited by the triangle inequality of the metric used.[11]

Since these sets are defined over a conceptual similarity relation (relative to a sentence relative to language and a probability function), these might be called conceptual similarity spheres and bear *some* structural resemblance to the similarity spheres used for counterfactual semantics by Lewis (1973), which might already be suspected when glancing at Figure 2. Most notable differences are that (a) the subjects of discourse here aren't possible worlds at all and instead sentences of a language, that (b) two sentences may inhabit the exact same location, that is, have the same conceptual role, whereas two possible worlds are distinct from one another and (c) that this is not an objective ordering of any kind and instead a purely subjective ordering based on an agent's probability function. Given such a probability function, however, the ordering is rather fix.

---

classes.

10. Although, of course, there may be distinct sentences with the exact same conceptual role, such that S_p^1 may include additional sentences, which is Field's point.

11. $\mathfrak{d}(q,r) \leqslant \mathfrak{d}(q,p) + \mathfrak{d}(r,p)$.

It would be interesting to investigate under which permutations of different measures the ordering stays invariant.

In short, through some simple structural features, a probability function yields a deep comparative notion of conceptual role of sentences and, one could presume, potentially some insights about meaning.

The next section develops a simple example to see how this notion of conceptual role is applied.

2.3 A Toy Example: The Lottery

To use a familiar example, let's consider a standard fair n-ticket lottery scenario in which an agents has beliefs about a very narrow set of sentences $\{t_1, \ldots, t_n\}$, where t_i is the sentence that the ticket number i is the winning ticket. Assume a probability distribution of someone who is aware of the real chances involved, i.e. that the sentences describe disjoint events, such that $\Pr(t_1) = \cdots = \Pr(t_n) = \frac{1}{n}$ and, more importantly

$$\Pr(t_i \mid t_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

We might, in this case, set $\mathfrak{c}(t_i, t_j) = \Pr(t_i \mid t_j)$, set $\vec{\mathrm{t_i}} = (\mathfrak{c}(t_i, t_1), \ldots, \mathfrak{c}(t_i, t_n))$, with $\vec{\mathrm{t_j}}$ accordingly and so a euclidean distances measure gives

$$\mathfrak{d}(\vec{\mathrm{t_i}}, \vec{\mathrm{t_j}}) = \left(\sum_{k=1}^{n} (t_{i_k} - t_{j_k})^2 \right)^{1/2} = \sqrt{2},$$

if $i \neq j$.

Hence, for any fixed t_i there are only two corresponding similarity spheres:[12] $S^1_{t_i} = \{t_i\}$ and $S^2_{t_i}$, which contains all other sentences.

This result isn't particularly exciting, but it serves well enough to indicate the spatial interpretation of conceptual role: Given no further information, t_1 and t_{10} are just as close in meaning as t_2 and t_{77}. It is easy to see that when adding additional constraints, such as that you yourself hold the ticket number 25 and adding the sentence w: "I win the lottery!", the previously strict symmetry is broken and t_{25} takes a spatially distinguished position, since $\Pr(t_{25} \mid w) = 1$, but $\Pr(t_l \mid w) = 0$ for all other l.

12. No matter how we chose the similarity function $\mathfrak{f}$, as long as it is monotonously decreasing, as required above.

2.4 Choosing a proper measure

As mentioned, choosing the right measure to find what exact quantity gets assigned to the axes of our spatial representation isn't perfectly straightforward. Generally, the idea is to use a measure which represents the evidential relation of two sentences p and q. This is closely related to what is usually called degree of confirmation[13]. Here are two very common measures, taken from Fitelson's extensive discussion:

$$\mathfrak{c}_d =_{df} \Pr(p \mid q) - \Pr(p)$$
$$\mathfrak{c}_l =_{df} \log(\frac{\Pr(p \mid q)}{\Pr(p)})$$

which both have the desirable properties that

$$\mathfrak{c}_{(\cdot)}(p,q) \begin{cases} > 0 & \mathit{if}\, \Pr(p \mid q) > \Pr(p), \\ < 0 & \mathit{if}\, \Pr(p \mid q) < \Pr(p), \\ = 0 & \mathit{if}\, \Pr(p \mid q) = \Pr(p). \end{cases}$$

That is, $\mathfrak{c}$ takes a zero value for Pr-independent p and q. This makes a spatial representation much more tractable, since irrelevant dimensions just don't have any spatial expansion at all and can be easily disregarded. This were not the case if, for example, it'd be just the conditional probability. The property is highly useful for the next section.

It is important to note that these two measures aren't ordinally equivalent (Fitelson 1999, 364), meaning that they potentially induce different orderings on $\mathcal{L}$, which is, of course, quite relevant to the present account.

A thorough treatment, then, would include a decision on which measure to use. Mentioning some options and desirable properties shall suffice for now.

3 Context Dependent Conceptual Role

The whole enterprise of a spatial interpretation of a sentence is somewhat of a holistic approach, in that it evaluates meaning *globally* for a fixed language and agent. It is, however, pretty uncontroversial that the lexically same sentence can have different intended meanings, depending on the circumstances of when it's uttered. Relative to different contexts, the same words may mean something different. Can the present account accommodate this?

13. cf. Fitelson (1999), Brössel (2013).

So far, there has been an implicit assumption that the conceptual role is represented relative to the probability function over the whole language. This needn't be the case, though, and it is mathematically of course perfectly legitimate to restrict the sentences which make up the dimensions of the spatial interpretation.

Stalnaker (1978) speaks of common ground amongst speakers in a conversation as establishing a context for utterances. This ground is determined by a context set, which consists in propositions assumed true for the purposes of a particular communication.

In a similar vein, one might adopt the notion of a context set — populated with sentences instead of propositions — as a means to restrict the spatial interpretation of conceptual role. Intuitively, this makes some sense: Asking for a meaning in a particular context, and having already accepted the assumptions about the connection between meaning and evidential relation described in this paper, it is quite natural to restrict the evidential relations that are interesting to us to to those that involve that particular context.

This needn't be just those sentences that correspond to propositions assumed true, but may be extended to sentences deemed *relevant* to the current interests. That relevance is commonly spelled out in terms of evidential relation makes determining a context seem circular, at first glance.[14] Assume a context already determined.

Intuitively, then, one way to work out a context is to "chip away" at the vector space and remove as many dimensions as needed such that only dimensions corresponding to a context set remain.

More formally, let V be a vector space over $\mathbb{R}$ with dimensions corresponding to sentences in a language $\mathcal{L}$ as above, with an appropriate measure $\mathfrak{c}$ chosen. Let $C \subseteq \mathcal{L}$ fix a context set of sentences that interest us, and let V_C be a vector space over $\mathbb{R}$ with dimensions corresponding to sentences in C. Then V_C is a subspace of V[15].

In a subspace like this, the distances between sentences can, of course, change notably compared to the surrounding space, since dimensions that made up the bulk of the distance could be disregarded. Each context, then, induces another family of orderings

14. It is not, though. Contexts can be generated from evidential relations through the spatial representation, as well. Since for two sets $C, D \subseteq \mathcal{L}$, $V_{C \cup D}$, $V_{C \cap D}$ and $V_{\mathcal{L} \setminus C}$ are subspaces of V, contexts can be combined at will. Given a sentence p, what are relevant other sentences? What could be considered a *total* context of p? Intuitively those sentences p stands in an evidential relation with, or is not probabilistically independent of. Depending on the measure $\mathfrak{c}$ chosen, these are exactly those sentences corresponding to the dimensions p takes a non-zero value (cf. Section 2.4). This gives a total context set for p. Accordingly, this can be generated with additional sentences. The intersection of both total contexts yields the common context.

15. Or, more correctly, there exists a unique isomorphism of V_C to a subspace of V that preserves assignment of sentences to axes and assigns 0 otherwise. For simplicity, I'll speak of subspaces, even though that isn't true in the strict sense.

on $\mathcal{L}$ by conceptual similarity.

The metaphor of a context as a subspace has some interesting advantages. It could be interesting, for example, to compare the change in conceptual role of some expression *across* contexts. In what — quantifiable — sense is an expression used differently when uttered in another context? If C and D are such context sets, two sentences p and $q \in \mathcal{L}$ (not necessarily in C or D) may differ wildly in distance in conceptual similarity given C vs. given D.

4 Inter-Personal Similarity of Conceptual Role

So far, the spatial interpretation of conceptual role has only been developed for a single language, with a single probability function. What is really interesting, though, is whether the account can be extended so as to include inter-speaker similarity in conceptual role. This is a much needed notion, as it strives to deal with an explication of why communication between two speakers actually works, although they, let's plausibly assume, do not mean *exactly* the same when using linguistic expressions.

The problem is, of course, quite notorious, and a hint at a possible solution here has to suffice.

Intuitively, though, the answer seems clear: two speakers use their expression *similarly enough* for communication to properly function, exact sameness in meaning is not required. How does that figure in this account? Is there a formal explication of this idea? A measure for the similarity in conceptual role of different speakers is needed.

In his 1979 paper, Field already probes this problem, and simply gives up:

> The problem is that different people have different subjective conditional probability functions: the machinery developed in this paper provides a natural account of what it is for two sentences or two terms *within the context of the same probability* function to have the same conceptual role; but I do not see any way to provide an account that is both clear and useful of what it is for terms or sentences in the contexts of *different* probability functions to have the same role. My own inclination is not to try to provide such an account, but the learn to live without the concept of inter-speaker synonymy, and all other concepts in terms of which inter-speaker synonymy could be defined.[16]

Field is not explicit about what his reasons are to reject the attainability of a clear and useful account given different probability functions to talk about sameness of conceptual

16. Field (1977, 398), his italics.

role, but one may suspect that it is for similar reasons as is described in the following, what one may call the *mapping problem* for spatial interpretations of inter-speaker similarity of conceptual role. Of course, Field speaks of *sameness* instead of similarity, but since sameness here is just the special case of maximal similarity, it inherits the problems posed for his account.

4.1 The Right Mapping Between Languages

The central problem is how to map the language of one agent to the other agent's language such that pairs of sentences will be assigned to a single dimension in the vector space. That is, when comparing the conceptual role of two sentences of two different agents, the aim is to find a way to represent both languages in the same vector space to make sense of the idea of a distance measure. But which sentences get paired up, as there might not be a one-to-one correspondence at all? If just a random sentence is chosen, all information is lost, one instead needs to find a pairing that mirrors most closely the *structure* of its evidential relations. This challenge can be formalized as follows:

Suppose two speakers with languages $\mathcal{L}$ and $\mathcal{L}'$ and probability functions Pr and Pr$'$ over their languages, respectively. The goal is to find a surjective[17] function $f : \mathcal{L} \to \mathcal{L}'$ such that two sentences $p \in \mathcal{L}$ and $p' \in \mathcal{L}'$ can be compared with respect to the measures of evidential relation $\mathfrak{c}$ and $\mathfrak{c}'$ to other sentences $c_i \in \mathcal{L}$ and $f(c_i) \in \mathcal{L}'$. If we have such a function, a vector for p is determined by $\vec{\mathrm{p}} = (\mathfrak{c}(p, c_i), \ldots, \mathfrak{c}(p, c_n))$ as above and for p' by $\vec{\mathrm{p}}' = (\mathfrak{c}'(p', f(c_i)), \ldots, \mathfrak{c}'(p', f(c_n)))$. The distance is then calculated as usual: $\mathfrak{d}(\vec{\mathrm{p}}, \vec{\mathrm{p}}') = \left(\sum_{i=1}^{n} (p_i - p_i')^2\right)^{1/2}$.

The simplest case assumes two agents with probability functions over the exact same language. In this case is $\mathcal{L} = \mathcal{L}'$ and the function $f : \mathcal{L} \to \mathcal{L}'$ is just the identity function, and we are done, it seems like.

But wait a minute: The talk of identity of sentencs has been quite careless so far. What does it mean that $\mathcal{L} = \mathcal{L}'$? The elements are identical. The criterion for sentential identity is lexical identity, such that the sentences "There are two cats on the mat" and "There are 2 cats on the mat" are *not* identical to one another, although they *mean* the same, let's assume. On the other hand, two lexically identical sentences may mean something different relative to two agent's probability functions. In fact, two agents may assign wildly different conceptual roles to lexically identical sentences. Using the lexically identical sentences as fixpoints to merge both speaker's spatial representations of conceptual role as described above, then, seems to just beg the question, or at best

17. We assume here for simplicity of notation that $|\mathcal{L}| \geqslant |\mathcal{L}'|$.

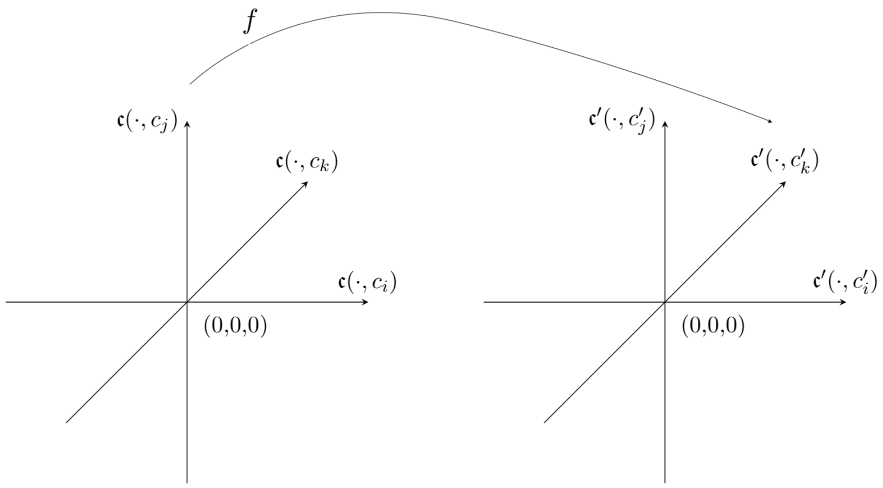

Figure 3: The *mapping problem*: How can the elements of $\mathcal{L}$ (left) best be mapped onto the elements of $\mathcal{L}'$ (right)? How can one make sense of both languages in a single vector space?

rely on an empirical assumption that most probably, the lexically same sentence will mean about the same. But this was the *goal* all along: find a measure for semantic similarity. This kind of arbitrariness is unsatisfying for a formal treatment, showing the need for a more general account that, when applied to a case like the simplest one with identical languages, yields in *normal* cases just the mapping of sentences to lexically identical ones, while allowing for mappings to other sentences as well, depending on the probability function.

What, then, should determine $f(c_i)$? Intuitively, the sentence of $\mathcal{L}'$ *closest in meaning* to c_i. In normal circumstances, this is quite often the lexically identical sentence, if available. In a general account, this might be any sentence of $\mathcal{L}'$.[18] Isn't this approach even *more* question-begging than the last? After all, how can semantic similarity be quantified if not through spatial representation? This is a valid concern, however, it's important to distinguish between a *definition* of f, which may be circular, and a *criterion* to actually determine $f(c_i)$. The former might not yield an answer to how to do the latter, which can be a non-trivial task.

How to define f now? This is just a formalisation of the intuitive idea above.

18. The usefulness of this criterion varies considerably with the commensurability of the languages. If $\mathcal{L}$ is a full-blown no-expenses-spared propositional language akin to a natural language, and $\mathcal{L}'$ is $\{p, q\}$, there is not much sense in asking about semantic similarity across languages. Lets assume somewhat comparable languages.

We let $\mathcal{L}'$ be $\leqslant_{c_i}$-ordered like in Section 2.2. Then $f(c_i)$ is a maximal element of $\mathcal{L}'$.[19] And that's it.

To summarize: in order to determine the similarity in conceptual role of two sentences p and p' of different speakers, we calculate the distance $\mathfrak{d}(p, p')$, for which we need a mapping $f(c_i) = c'_i$ to determine the terms $\mathfrak{c}(p, c_i) - \mathfrak{c}(p', c'_i)$. To determine $f(c_i)$, we assume $\mathcal{L}'$ to be $\leqslant_{c_i}$-ordered, which just means that we ordered all sentences in $\mathcal{L}'$ by conceptual similarity to $c_i (\in \mathcal{L})$. In order to do this, we need to calculate the distance $\mathfrak{d}(c_i, d')$ for all sentences $d' \in \mathcal{L}'$, for which we need a mapping $f(e_j) = e'_j$ to calculate the terms $\mathfrak{c}(c_i, e_j) - \mathfrak{c}(d', e'_j)$, and so on.

The circularity needn't be a devastating thing, though, as there might exist a recursive algorithm that, given some reasonable assumption, creates a mapping such that the circular definition is satisfied. Such an algorithm is out of the scope of this paper, of course, but here goes a glimpse: Build a mapping from the ground up. Assume as a reasonable termination condition some fixpoints of f, that is, some such mappings as given. Let's call this set $F \subseteq \mathcal{L}'$. Then, incrementally increase the mappings by adding one sentence p of $\mathcal{L}$ at a time. That is, determine the closest sentence in $\mathcal{L}'$ in the subspace V_F. Add that mapping to F, and repeat, until every sentence of $\mathcal{L}$ is mapped to $\mathcal{L}'$[20].

The advantage of this approach is its generality: Let's assume, like in the beginning, the simplest cast of two speakers with $\mathcal{L} = \mathcal{L}'$. Now assume a normal case, in which most of the lexically identical sentences are closest in conceptual role, too, and the whole mapping is easily built.

The generality extends to more difficult cases, too, and enables to treat cases of quite dissimilar speaker-languages. The quality of the "translation" might drop steadily, though, when the intersection of both languages becomes increasingly small.

4.2 A Toy Example: Pastries

Let's illustrate with a simple example. Assume two speakers, both speaking the same natural language. They share a healthy appetite for a dough-based regional pastry. Their languages are almost lexically identical, let's assume, with the slight contingent difference that speaker α refers to the pastry as 'Berliner' while speaker α' never even heard of the term and uses 'Krapfen' instead. Let's call the intersection of both languages

19. Maximal element because the ordering is not antisymmetric, meaning that there are potentially multiple maximal elements. For this case, we need to chose one. Maybe *then* a lexical criterion could be applied, comparing the lexical similarity e.g. by taking the Levenshtein-distance.

20. Reasonable constraints apply, of course. It could be, for example, that there isn't a semantically most similar sentence for each sentence in $\mathcal{L}$.

$\mathcal{L} \cap \mathcal{L}' = F$ and generously assume lexically identical sentences to also be semantically most similar, as described above. Consequently, every element of $\mathcal{L} \setminus \mathcal{L}'$ has the word 'Berliner' in it, and 'Krapfen' *vice versa.*

With this setup, we try and find the semantically most similar sentence of $\mathcal{L}'$ to $s =$ 'Ich mag Berliner!' of $\mathcal{L}$. According to the definition, we ought to find the maximal element of $\mathcal{L}$ under $\leqslant_s$, for which we need to calculate $\mathfrak{d}(\vec{s}, \vec{l'})$ for each $l' \in \mathcal{L}'$. But we can't — there is no mapping for $\mathcal{L} \setminus \mathcal{L}'$. There is one, however, for F — the identity function. We now can approximate the most similar element of $\mathcal{L}' \setminus \mathcal{L}$ to each element of $\mathcal{L} \setminus \mathcal{L}'$ by using the subspace V_F, expanding the subspace in each iteration to include an axis for the element added in the previous iteration. At last, unsurprisingly, we determine the closest element to 'Ich mag Berliner!' as 'Ich mag Krapfen!'.

5 Conclusions, Objections, and Future Research

I hope to have given a somewhat convincing glance on what could be a useful approach to spell out conceptual similarity. Of course, it's peacemeal work at best. I tried to communicate the few core ideas quite succinctly, while only hinting at possible applications. In this concluding section, I'd like to discuss some salient objections to the account and briefly hint at possible responses for it. Finally, I suggest some of the many still lacking parts of this account.

It might raise a few suspicions to use a normative epistemological concept, rational degree of belief in form of a subjective probability function, to explicate semantic meaning. In particular, it is not clear whether the resulting account is normative. Consider: A mistake in reasoning, as for example a violation of the probability axioms, seems to be a normative mistake in the epistemological category. On this account, it would also influence directly the *meaning* of the involved sentences, as they shift from what it *should* have been epistemologically to a locally irrational state. Does this normativity transfer, so to speak, on the semantics of sentences? This seems, on the face of it, like two separate issues that best stay separated. A response for this account could grant this connection, and maybe deny that it is such a bad thing. If meaning is spelled out in the way described here, and an agent's understanding of a language is such that it entails inconsistent degrees of belief, then yes, it may be completely fine to assume that something of that understanding goes amiss.

A related concern is that since learning about anything in the Bayesian framework always involves shifts in degree of belief, on this account learning also involves shifts in meaning. Maybe it is not so helpful to tie degrees of belief to meaning at all! The concern is easily dispelled, though, as it rests on a misunderstanding of the present account. The meaning of a sentence is not represented by a location in the space, but instead the degree of similarity in meaning of two sentences is represented by their distance in this space. Maybe the *meaning of a sentence simpliciter* could be assessed by analysing the structure of its position in space relative to all other elements of the language, but that is a much broader question. The distance between two meanings might well stay constant over a Bayesian update, since those degrees of belief get updated, too. Should there be a change in distance it might still prove plausible, however, for it of course *might* change one's understanding of a concept when one is presented with new evidence, and additionally open up a way to incorporate diachronic changes in meaning to the present account.

The distance relation as a metric for similarity has already been criticised by Tversky back in 1977[21]. It is not plausible and even runs counter to empirical findings, Tversky argues, that judgments about similarity relations do in general exhibit the symmetry requirement of a metric. For example, Tel Aviv may be judged more similar to New York than New York is judged similar to Tel Aviv in empirical studies. A proponent of the current account could answer that judgments do not need to reflect actual representation, or, if there is surmounting evidence that it does, retreat to the position that this account was never meant to be descriptive at all. A proper response, however, distinguishes between different contextual features of the similarity judgments. As Gärdenfors points out, comparing Tel Aviv to New York may highlight different features for comparison than comparing New York to Tel Aviv. Consequently, one may analyse these judgments as taking place in different contexts, which may very well yield different spatial distances, as described in Section 3.

One may also worry about the seeming arbitrariness of weights of axes: each sentence is weighted equally when calculating the distance measure. Intuitively, though, there are huge differences between important and negligible sentences, yet on this account they all matter the same amount. I agree that this is a viable criticism, and would very much like to see a compelling account of this importance, to be able to figure it into the present one. Apart from suggesting that one could weight the different axis accordingly, I have no solution at present.

Leitgeb (2008) constructs an impossibility result for the very concept of meaning similarity, concluding that either its doomed or we'd have to give up a substantial bit of compositionality. In lack of a proper treatment of this topic due to time and space constraints, let me just list a few reasons why this result could be argued to not apply

21. cf. Tversky (1977), also Gärdenfors (2004, 112) for a discussion

here: (i) There is no compositionality — as of yet — alluded to at all in this account. Granted, that should be changed, so this is not a convincing response. (ii) The languages that figure in this account might be restricted such as to exclude those problematic for Leitgeb's result, but since this is a general approach, one might want to instead explicitly include them. (iii) The account here does not feature a binary similarity relation at all, as in 'p is meaning similar to q', and instead merely yields a comparative 3-place relation as in 'p is more similar to q than to r'. If one introduces such a notion, it's still open to argue that (iv) connectedness does not hold in general. Since the present space is sparse with about as many vectors as dimensions, distances between vectors may be such that there are clusters of sentences, without distances less than an ϵ to the next cluster for any element. Concepts are represented as vectors or points, but not as adjoining regions, where connectedness is much more plausible. This response is, as noted, just a sketch and in dire need of a more thorough treatment of more work.

The above discussion sheds light on a major shortcoming on the presentation of the account so far: No constraints imposed on meaning by logical relations between sentences have been considered so far. If one takes the probability axioms seriously, that is, how logical relation between sentences figure into the spatial relations, there are very probably interesting phenomena to discover. This could have been fleshed out with some examples, an idea that immediately springs to mind is the comparison of rational / non-rational credence functions in the Linda scenario.

So far, I have spoken interchangeably of sentences and linguistic expressions, as if those two terms were extensionally equivalent. But of course there are much more linguistic expressions in need of a semantic treatment, notably names or constants and predicates, as Field (1977, 396) suggests. He also sketches an idea of what this might amount to in the case of sameness of conceptual role, which might also give an idea of how to design it for similarity of conceptual role. I exclude a treatment from this paper, and merely note that it could be very interesting to expand in this direction.

Churchland (1998) proposes a different way for this account to tackle the problem of inter-speaker similarity: Instead of trying to explicitly map one axis to another, one could instead look for structural similarities of clusters of locations in the semantic space. That is, two speakers would mean something very similar with a sentence if, in both speaker's spaces, the sentence appears in similar geometric patterns with other sentences. This complex structural analysis might actually yield interesting results.

References

Brössel, Peter. 2013. “The Problem of Measure Sensitivity Redux.” *Philosophy of Science* 80 (3): 378–397.

Churchland, Paul M. 1998. “Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered.” *The Journal of Philosophy* 95 (1): 5–32.

Erk, Katrin, and Sebastian Padó. 2008. “A structured vector space model for word meaning in context.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 897–906. Association for Computational Linguistics.

Field, Hartry. 1977. “Logic, Meaning, and Conceptual Role.” *Journal of Philosophy* 74 (July): 379–409.

Fitelson, Branden. 1999. “The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity.” *Philosophy of Science* 66 (3): 378.

Fodor, Jerry A., and Ernest Lepore. 1999. “All at Sea in Semantic Space: Churchland on Meaning Similarity.” *Journal of Philosophy* 96 (8): 381–403.

Gärdenfors, Peter. 2004. *Conceptual spaces: The geometry of thought.* The MIT Press.

———. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces.* MIT Press.

Goodman, Nelson. 1972. *Problems and Projects.* Indianapolis: Bobbs-Merrill.

Leitgeb, Hannes. 2008. “An Impossibility Result on Semantic Resemblance.” *Dialectica* 62 (3): 293–306.

———. 2013. “Reducing Belief Simpliciter to Degrees of Belief.” *Annals of Pure and Applied Logic* 164 (12): 1338–1389.

Leitgeb, Hannes, and Richard Pettigrew. 2010. “An Objective Justification of Bayesianism I: Measuring Inaccuracy.” *Philosophy of Science* 77 (2): 201–235.

Lewis, David K. 1973. *Counterfactuals.* Blackwell.

Sahlgren, Magnus. 2006. “The Word-space model.” PhD diss., University of Stockholm (Sweden).

Stalnaker, Robert. 1978. “Assertion.” *Syntax and Semantics (New York Academic Press)* 9:315–332.

Tiffany, Evan. 1999. “Semantics San Diego Style.” *Journal of Philosophy* 96 (8): 416.

Tversky, Amos. 1977. “Features of similarity.” *Psychological Review* (US) 84 (4): 327–352.