PENULTIMATE DRAFT
(Final version published in: *Pragmatics & Cognition,* (2019) Vol. 26:2/3, 239–266.)

The Liar Paradox in the Predictive Mind

Christian Michel
The University of Edinburgh

Abstract

Most discussions frame the Liar Paradox as a formal logical-linguistic puzzle. Attempts to resolve the paradox have focused very little so far on aspects of cognitive psychology and processing, because semantic and cognitive-psychological issues are generally assumed to be disjunct. I provide a motivation and carry out a cognitive-computational treatment of the liar paradox based on a cognitive-computational model of language and conceptual knowledge within the *Predictive Processing* (PP) framework. I suggest that the paradox arises as a failure of synchronization between two ways of generating the liar situation in two different (idealized) PP sub-models, one corresponding to language processing and the other to the processing of meaning and world-knowledge. In this way, I put forward the claim that the liar sentence is meaningless but has an air of meaningfulness. I address the possible objection that the proposal violates the *Principle of Unrestricted Compositionality*, which purportedly regulates the conceptual competence of thinkers.

Keywords: Liar Paradox, Predictive Processing, Semantic Paradox, Unrestricted Compositionality, Yablo Paradox

1. Introduction

Consider the following version of the liar paradox. Mary says, *I am now lying*. If what Mary says is true, she must speak falsely, and if she speaks falsely, she must be telling the truth. So, we have a contradiction, or we must assume that what Mary says is neither true nor false. The extreme simplicity of its statement and the concern that the paradox might reveal fundamental inconsistencies in our basic intuition about logic, language and the concept of truth (see Beall, Glanzberg & Ripley 2019: section 3) have led to an enormous number of attempts to resolve it.

Many of the proposals to resolve the paradox try to argue that the liar sentence is meaningless. If the liar sentence is meaningless, then any reasoning with it is pointless. Therefore, the paradox is blocked. An alternative approach to resolving the paradox is to modify classical logic and semantics. For instance, one might declare that the liar sentence is truth-valueless, i.e., it falls into a gap between true and false (paracomplete approaches, e.g., van Fraassen 1970, Kripke 1975) or has a third truth-value; or--quite counter-intuitively--that it is both true and false (paraconsistent approaches, e.g., Priest 1984, 2006). Barwise and Etchemendy (1987) proposed another account based on non-classical logic. They construct a semantics based on non-standard set-theory, namely, *hyperset theory*, in which sets can contain themselves.

The first sort of solution, via the meaninglessness of the liar sentence, is often unsatisfying because of either a lack of independent motivation for declaring the liar sentences meaningless or because the solution comes at the price of highly counterintuitive consequences. For instance, Tarski suggests that the truth predicate is stratified and can only be applied from a meta-language level to an object language level. However, in natural language truth does not seem to work that way; the condition that no language can contain its own truth predicate seems too restrictive. There is, for example, no problem with an honest person saying *I am always telling the truth*. Also, Etchemendy and Barwise's solution leads to strange implications. Their account is based on hyperset theory, which presupposes a notion of *self-containing sets* (hypersets are no longer conceptualised as collections of things, but as directed graphs allowing for cyclical relations). This has the weird implication that we cannot make statements about the entire world (1987:174).

On the other hand, to easily give up classical logic--as the second main type of solution suggests--does not seem to be a good idea either, given its success as both common-sense logic and the logic of science and mathematics (see also Williamson 1996, Leitgeb 2007: 283). It seems astonishing that logic would have to be replaced in the domain in which the liar is situated: a quite trivial domain of everyday life (with lying and truth-saying people). Solutions with gaps or third truth values also suffer from the 'revenge problem'. One can reformulate the liar sentence in a way that makes it reappear in a different form. For instance, if Mary says *I am either lying or saying something that is neither true nor false*, then this sentence is also subject to the same sort of paradoxical reasoning with regard to the question of whether it is true or false or neither true nor false.

The discussions and analyses of the liar paradox in the literature along the above lines implicitly assume that it is a formal logical-linguistic problem. Logic, language, and concepts like TRUTH are further treated implicitly, like abstract, mind-independent objects and cognitive-psychology has played no significant role in any influential solution attempt. Maybe Leitgeb's (2007) requirement that "[...] ultimately every successful philosophical theory of truth has to stand the test of formalization [...]" (p. 276) best summarises the formalistic tradition in regard to resolving the paradox. There are some exceptions, though, with regard to the treatment of the liar paradox as a formal *semantic* paradox. Most notably, Martinich (1983) suggested that the paradox is *pragmatically* based (e.g., p. 64) in the sense that the speaker fails to execute a proper assertoric speech act when uttering a liar sentence. However, Martinich does not explain on independent grounds *why* the speech act fails (beyond the fact that it produces a paradox) and does not provide any discussion on cognitive-computational processes. Also, there are some recent psychological approaches to the liar paradox in the context of experimental psychology. However, those treatments are purely descriptive of how participants actually judge liar sentences in terms of truth value assignments (e.g., Elqayam 2006. Ripley 2016) and they do not pretend to 'resolve' the paradox, nor do they discuss cognitive-computational processes. Rips (1989) is an exception and proposes a "computational model" (p. 90) for how we reason with liar-type sentences. According to him, we carry out inference processes that are formalizable by a Gentzen-type natural deduction framework. Kearns (2007) follows Martinich's speech act approach but formalizes it into a system of "illocutionary logic". Rips and Kearns, therefore, remain in the formalists' territory.

I do not deny that formalisations are desirable and useful. However, one could suspect that this requirement has some cognitive origin itself. Our 'desire' to formalise could be some unconscious norm or bias, which strongly penetrates our scientific thinking. Furthermore, empirical findings from cognitive neuroscience support the idea that the way we think is shaped by the specific architecture of the brain, the body, and the environment. Thinking is not (only) a process carried out via language-like syntactic operations with amodal symbols (see Fodor''s (1979) "Language of thought"). Rather, thought relies on re-enacting sensorimotor states involving the activation of modality-specific representations in the brain (e.g., Hoenig, Sim, Bochev, Herrnberger & Kiefer 2008, van Dam, van Dongen, Bekkering & Rueschemeyer 2012, van Dam, van Dijk, Bekkering and Rueschemeyer 2012). But if this view of cognition is in the right ball-park, then there might be fundamental doubts about the possibility of full formalisations of language, logic, and concepts, because formalisations rely on amodal symbol systems.

Therefore, this paper reconsiders the liar paradox by providing a cognitive-computational treatment. I suggest that this approach allows for independently motivating the meaninglessness of the liar sentence and for retaining classical logic. It explains the paradoxical feel of the liar and avoids other counter-intuitive implications. My plan is as follows. In Section 2, I will provide a motivation for a cognitive-psychological treatment of the liar paradox (as opposed to the prevailing "formalist" treatments). Then I will briefly lay out a cognitive-computational model for language and concepts within the relatively new, but already well-established*, Predictive Processing* (PP) paradigm (Section 3). In Section 4, I will explain how the paradox might arise cognitively within the PP model and how we could motivate the meaninglessness of the liar sentences. In Section 5, I will discuss a critical possible objection involving the *Principle of Unrestricted Compositionality* of concepts, and in Section 6 I will conclude.

2. Motivating a cognitive-psychological treatment of the liar paradox

Why might it be worthwhile to tackle the liar paradox from a cognitive-computational perspective, as opposed to taking it as a formal linguistic-logical puzzle with a solution to be sought inside a formalised system? One answer is that the cognitive access to the liar situation and resulting paradox--i.e., the grasping and appreciating them--is mediated by language and cognitive processing. We are confronted with a liar situation in which a particular sentence, the liar sentence, is uttered. Only under the assumption that the liar sentence expresses immediately, directly, and transparently a mind-independent proposition (i.e., some abstract entity) could we avoid the intermediate step of some cognitive and language processing. However, such a view is not very plausible. It might be that, ultimately, the paradox is inherent in the way the mind-independent world is structured and we merely 'discover' the paradox. But it seems more plausible that the paradox arises from how we represent the world conceptually and linguistically and how we reason. If there is cognitive mediation, then there is at least a possibility--that we should not easily ignore--that the paradox arises because something goes awry with the cognitive processing. That much, so it seems, must even be admitted by die-hard referentialists (e.g., Fodor & Pylyshyn 2015). For referentialists, semantics depends only on the (mind-independent) referents of concepts, which are established, for instance, via some causal connection between the referent and the mental representations of those concepts. For referentialists, cognitive content is no more than some "[...] aura of associations, attitudes, feelings, beliefs, quasi-beliefs, recollections, expectations [...]" (Fodor & Pylyshyn 2015: chapter 2 and p. 146) associated with a referent.

A second way to justify why looking at the liar paradox from a cognitive angle might be valuable is as follows. There is already a huge number of solution proposals available. Those solutions all produce some counter-intuitive consequences. Let us assume that there is a solution, and whatever the solution is, we have to pay some price in terms of counter-intuition. This is not implausible; consider, for example, the fact that one of the most successful scientific theories in history in terms of experimental support, quantum physics, is highly counter-intuitive. Now 'intuition' is very much a cognitive-psychological notion. Why don't we turn, therefore, to cognition to try to adjudicate a resolution to the liar based on some understanding of how the counter-intuition arises? Here we would not reject an existing formalistic solution, but instead would use a cognitive approach as a complementary tool for adjudication.

A third possible answer goes far beyond this minimal concession, which even referentialists could make, and points to the paradigm of embodied cognition. According to this paradigm, conceptual representations, meaning and thought are grounded in the sensorimotor experience with the world and are hence shaped by the specific characteristics of our body and brain. Even iconic examples of formalistic disciplines, like formal logic, mathematics, and grammar, are grounded in cognition and not a world of mind-independent objects. Thought cannot be reduced to the syntactical processing of amodal symbols. Let me provide three examples to illustrate the idea that formal symbol systems are cognitively grounded.

Firstly, Dutilh Novaes (2012) suggests that formal languages, like logics or mathematically expressed theories in physics, are cultural artefacts. Their function is to de-bias thinking, especially in scientific contexts. Since the work of Tversky and Kahneman (1974), it has been well-known that the way humans actually think is strongly biased and violates many principles of rationality. Reasoning with the help of formal languages, according to Dutilh Novaes, is externalised "sensorimotor engagement" (Dutilh Novaes 2012: 162) with symbols, applying certain rules mechanically without taking into account the meaning of the symbols. One way to spell out what it means to be 'formal' is via the notion of *de-semantification*. De-semantification allows us to "switch off" cognitive content that is automatically associated with concepts ("semantic activation") and hence to suppress biases, for instance, in the form of prior beliefs (Dutilh Novaes 2012:206–207). The mental manipulation of symbols is a "pushing around" of those and involves modality-specific (sensorimotor) areas of the brain. Dutilh Novaes further observes that writing in a formal language is not an a posteriori expression or description of thoughts and cognitive phenomena. Instead, the formalism is a vehicle of thought; languages are cognitive tools that enhance cognition. Furthermore, Dutilh Novaes takes an explicit position with regard to the object of formalization. It is not a portion of the world, but theories or concepts that we have. Hence what is formalised are *intensional* objects (e.g., Dutilh Novaes 2012: 224). For example, Peano arithmetic is a formalisation not of a series of numbers but our theory of it. With this characterisation of the object of formalisation, we can avoid the question of whether numbers or logics exist independently of the mind or are merely mental constructs. All we are theorising about are the ideas and notions of numbers that humans have, not numbers themselves. Dutilh Novaes' position allows us to remain agnostic with regard to difficult ontological issues.

As a second example, take Lakoff & Núñez (2000). They suggest that mathematics and logic are not disciplines constituted by abstract, mind-independent objects and truths. Rather, all mathematical concepts are embodied and cultural artefacts, i.e., "a product of the human mind" (Lakoff & Núñez 2000:9). Central to Lakoff and Núñez' account is the idea of "image schemas", which can be understood as a basic modality-specific mental representation. Visual-spatial image schemas correspond, for instance, to concepts expressed by prepositions like *in*, *on*, *at* or *above*. The "Container Schema", (which is a *gestalt* consisting of some boundary, some inside and outside) is central to mathematics (on which, e.g., set theory is grounded). They are both conceptual and perceptual (Lakoff & Núñez 2000: 31). Take a branch of mathematics, arithmetic, as an example. According to Lakoff and Núñez, we conceptualise the objects and principles of arithmetic in the form of "conceptual metaphors". One such metaphor is ARITHMETIC IS OBJECT COLLECTION. We understand the concepts and principles in the "target domain" ARITHMETIC by transferring them from the "source domain" OBJECT COLLECTION. The source concepts stem from experience, like taking away or adding objects from collections. From correlations between manipulating object collections (which are sensorimotor operations) and arithmetic operations arise neural connections that constitute a "conceptual metaphor at the neural level" (Lakoff & Núñez 2000: 55). Image schema like the IN schema have modality-specific "spatial logics" on which amodal, formal symbolic logic is grounded. For instance, the law of modus ponens is grounded in the following way. Given two container schemas A and B, and an object X, if X is in A and A is in B, then X is in B. But it is not only neural visual-spatial areas, but also the motor control system, that are involved in mathematical conceptual thought (Lakoff & Núñez 2000: 34–35).

As a last example, take Langacker's *Cognitive Grammar* which applies the idea of the embodiment of cognition to grammar. Contrary to the traditional view, Langacker (e.g., 2008) denies a formalistic view of grammar, the idea that it can be represented exhaustively in the form of an amodal symbolic system:

> The picture that emerges belies the prevailing view of grammar as an autonomous formal system. Not only is it meaningful, it also reflects our basic experience of moving, perceiving, and acting on the world. (Langacker 2008: 4)

Langacker's surprising claim is that grammar is meaningful, and that grammar and lexicon differ only in degree, not nature. He motivates this view via examples of "partially schematic units", which can be classified neither as paradigmatically lexical in nature, nor as pertaining exclusively to grammar. The following linguistic schema: *Vs* X IN THE *Nb* (where *Vs* is a verb meaning 'striking' like *hit*, *kick*, *strike*, or *poke* and *Nb* is a body-part noun like *shin*, *back*, *face*, *eye*, or *knee*) is an example of such a partially schematic unit (Langacker 2008: 20). Now, according to Langacker, (e.g., 2008: 5) grammar is meaningful because it is "symbolic", much like lexicon. A "symbol" is a pairing of a "semantic representation" (which is a complex embodied conceptualisation, much along the lines of Lakoff & Núñez (2000)) and a "phonological representation" (including as well gestures and orthographic representations). Grammar and lexicon form a gradation in terms of specificity/schematicity. Grammatical symbolic units are just more schematic than lexical ones, but they are not different in nature.

All three examples show how cognition takes centre-stage when analysing formal systems. If they are in the right ball-park we should consider formal systems like logic, language or mathematics not as mind-independent abstract amodal symbol systems but as grounded in modality-specific representations. Maybe the deadlock in resolving the liar paradox via the formalistic approach is due to neglecting the embodied cognitive basis of formal systems. Therefore, it seems at least worthwhile to explore a cognitive approach to the semantic paradoxes and see where it leads us. Such an approach requires a specific cognitive computational model and an account of concepts and language, which I will outline in the next section.

But before that let me address the following possible objection to an embodied cognitive treatment of matters related to logic. Normally, the nature of logic is expressed as a dichotomy: it is either descriptive of or a norm for rational thought. But, for instance, the principles of classical logic are not descriptive of how in fact we reason, as the literature on cognitive biases teaches us (see e.g., Tversky & Kahneman 1974, for the classical paper). That leaves us with option two. But if logic is a mind-independent norm then how we actually think seems irrelevant to a treatment of the liar paradox. I suggest denying the dichotomy and taking logic to be both descriptive and normative. It is descriptive of what we think the norm of rational thinking should be.[1] Following the idea of Dutilh Novaes regarding what the objects of formalisations are, the analysis here will be of an intensional object (our understanding and ideas of the rational norm), rather than the "norms themselves" (if they exist at all in the Platonic heaven). In other words, we can go agnostic about whether the norm exists independently of the mind, without falling into an unpalatable idealism or solipsism.

[1] As we will see later in section 5.2., the idea of such an internal norm can be spelled out naturally under the framework of a hierarchical generative model.

3. Predictive Processing and a dual account of language and conceptual knowledge

3.1. The Predictive Processing framework

According to Predictive Processing (PP) (see Clark 2013, 2016, Hohwy 2013, Friston 2010), the brain is an embodied, multi-level prediction machine that continually anticipates its sensory input, relying on a mental prediction model. The PP model structure has the form of a generative probabilistic model in which approximate Bayesian inference is carried out (e.g., Clark 2013: 188–189, Hohwy 2013: 15–39). The PP model has a hierarchical structure and represents prior knowledge on many levels of abstraction (e.g., Clark 2013: 25, Lupyan & Clark 2015). In the top-down prediction cascade, the predictions of higher-level layers serve as priors for the lower-level predictions and, in this way, constrain the hypothesis space on the lower level. The main feature specific to the PP story is that the computations in the brain are driven by prediction error-minimisation and, therefore, the error signals play a central role. The predictions are compared to the actual sensory input in so-called *error units*, and the *residual error* of the predictions is calculated. It is then the error signals that flow laterally and upwards in the hierarchical network and may lead to updates of the model at different levels of the hierarchy.[2] The PP model is constantly adjusted in order to converge towards a version that minimises the overall average prediction error in the long run. There are two fundamental and interrelated ways to minimise the prediction error: firstly, by updating the internal model to fit the predictions to the sensory flow; and secondly, by generating actions to fit the sensory flow to the predictions (*active inference*), i.e., the brain with its body can change the world to fit its prediction. In this way, PP brings action, perception, and cognition into a unified framework under a unified (embodied) cognition paradigm. The prediction error minimising process is supplemented by a mechanism of precision-weighting of the prediction errors (Clark 2016: 53–83). The brain needs to discriminate noise and useful signals because noise should not force an update of the model. The brain must, therefore, predict the reliability of the sensory input, assign weights to the error signals and thus determine the influence of the top-down predictions versus updates driven by the bottom-up error-signals of the model. An example from Clark (2013: 198) might serve to illustrate this

[2] The mechanism of processing only the error signal is also called *predictive coding*.

point. In a situation of thick fog, visual sensory information about the shape of an object is less reliable than, e.g., tactile or auditory information. In such a context, the precision-weighting mechanism predicts that the bottom-up visual signal has low precision. Bottom-up error signals related to the shape are then tuned down to avoid an update of the brain's prediction model, and the influence of other sensory modalities or top-down predictions increases.

While the PP model has a unified architecture (often represented as a hierarchical probabilistic graphical model), it is compatible with a certain functional modularity, where a module corresponds, roughly, to some domain. For instance, we can speak of a *sub-model* corresponding to folk-psychology in which concepts and knowledge are encoded relevant to predicting and explaining the behaviour of others based on desires and beliefs. Another example is folk-physics, a sub-model that encodes common-sense knowledge about how physical objects behave.[3] Those sub-models should not be seen as 'encapsulated'. Rather, they might arise out of hard-wired biases or be the consequence of the error-minimisation based optimisation of the PP model. Such domain-specific knowledge is still part of a single, large and highly interconnected overall PP model. This is important to keep in mind for the next section.

3.2. *The Language and Situated Simulation (LASS) Model*

The treatment of the liar that I propose is based on the idea that thought proceeds in a synchronised way in two sub-models (as qualified in the previous section), one for (formal) linguistic processing and the other for processing meaning and conceptual knowledge. This duality of sub-models should be taken as an idealization. There might be more than two sub-models involved in meaning and conceptual processing, for instance, a 'pragmatic' sub-model.[4] Such a pragmatic sub-model would encode the knowledge that allows the individual to derive the speaker's intentions and meanings. For simplicity, those other sub-models are grouped into the sub-model for meaning and the sub-model for conceptual processing. These two idealized sub-models rely on language-like linguistic representations and modality-specific representations, respectively. The paradox arises--I suggest--as a failure in the synchronisation of the

[3] See, e.g., Gerstenberg & Tenenbaum (2017) for a discussion of intuitive theories.

[4] Thanks to an anonymous reviewer for raising this point.

two modes of processing. Several authors (e.g., Paivio 1990, Simmons et al. 2008, Dell & Chang 2013) have proposed accounts of thought and language processing that make a dual distinction of representational types. For instance, Dell and Chang (2013) suggest a *Dual-Path* model for language production and understanding, with a dual structure consisting of coordinated processing on a "meaning pathway" and a "sequencing pathway" (e.g., Dell & Chang 2013: 4). I will adopt a version of the account of Simmons et al. (2008) (also Barsalou et al. 2008), however, with a specific PP-twist to it. The authors propose that conceptual processing relies on both language-like and modality-specific representations, in a way that depends on the context and cognitive task. According to their *LASS* (*Language and Situated Simulation*) model, in conceptual processing both the linguistic and simulation sub-systems become active. However, activations of the two types peak in different orders, depending on whether there is a linguistic or a perceptual cue: "Once a word is recognized, associated linguistic forms are generated as inferences, and as pointers to associated conceptual information" (Simmons et al. 2008: 107). Once the linguistic forms are associated, the system can pursue different strategies corresponding to different levels of conceptual depth. The linguistic form (or associated statistical information) might be sufficient for the purpose at hand; therefore, we might have a "shallow activation" of meaning. After the activation of the linguistic form, the brain might also activate modality-specific regions within itself, which generate a simulation of perceptual or motor mental states that would be active if it were to interact with the referent of the linguistic form. Those simulations are "situated" or context-specific and are often quick and automatic (less than 200 ms after word onset) and "deeper" conceptual representations, as opposed to the shallow, merely syntactic, ones (Simmons et al. 2008: 107). Different tasks imply a different mixture of the two representational forms. The linguistic system represents mere form, while the simulation system represents the "meaning". The authors stress that speaking of "two systems" is not to imply that there are two rigid modules; rather, it is a "simplification so that we can focus on mechanisms of interest" (Barsalou et al. 2008: 253). This corresponds to my idealization in the form of two sub-models mentioned at the beginning of this section.

3.3. The LASS model and Predictive Processing

The reason why I endorse the LASS model is that it is fully compatible with the Predictive Processing framework, and the PP approach provides further unity to the dual LASS account. The PP framework with its precision weighting mechanism also supplies a computational underpinning for the LASS model. Let me briefly develop those claims.

LASS and PP are compatible because we can treat the two sub-systems of LASS as sub-models in the overall PP model that we entertain. Language and its linguistic objects are further 'things in the world' that are modelled by the brain. The linguistic sub-system has a certain functional identity as a highly interconnected sub-model and encodes knowledge of a specific domain of things in the world, in this case, linguistic objects, like phonemes, words, and sentences. The PP model has an innumerable number of such sub-models, which are understood as closely interconnected parts of the model that track a certain world domain. For instance, when talking about moving objects, we can say that the sub-models corresponding to the two domains folk-physics and language are active and salient. The linguistic sub-model supplies the formal aspects of language, and the folk-physics sub-model provides the conceptual meaning.

There is also increasing evidence that language production and comprehension are predictive on all levels of the hierarchy of linguistic objects: phonemes, words, sentences, and discourse (see, e.g., Barsalou 2009: 1286, Pickering & Garrod 2013, Kutas et al. 2011, Kuperberg & Jaeger 2016, Allopenna et al. 1998, and also, Gagnepain et al. 2012). Therefore, the linguistic sub-system of the LASS model fits with the prediction-centred paradigm of the PP framework. PP can further provide a computational underpinning for the idea reflected in the LASS model that both sub-systems might receive different emphasis, or attention, at different moments. Clark (2016: 64–65) has suggested that the PP precision estimation mechanism can give control to different areas in the brain. For instance, visual information can be given priority over auditory information in the case that the estimated reliability of the latter is low. Similarly, I suggest that the mind can focus attention on the language sub-model or some other domain sub-model with non-linguistic conceptual content. In one case, one focuses more on formal-syntactical aspects during language comprehension (e.g., when drawing logical inferences content is irrelevant). The different words (as label-like linguistic objects) function as empty placeholders (in Dutilh Novaes' terms "de-semantisized objects") and computations run on the language sub-system only. But one can also focus more on conceptual content and lower the precision estimate for the language sub-system. This allows one to make sense, for example, of grammatically incorrect sentences, by prioritising the prediction of conceptual content, and not being too picky about grammaticality.

For convenience, throughout the remainder of the paper I will call the linguistic sub-model the *LSM* and the sub-model(s) of the domains relevant for a cognitive task the *WSM* (world sub-model). The LSM represents all formal, linguistic aspects (including statistic information about word co-occurrences, for instance). The WSM represents the meaning and conceptual (non-linguistic) knowledge. Again, this is a simplification, and I follow what the LASS authors have pointed out: there is no rigid modularity.

4. A cognitive approach to the semantic paradox

With a cognitive model for language and conceptual processing in hand, we now turn to a treatment of the liar paradox. The central idea is that we can conceive of a liar situation in two ways, one corresponding to a simulation/prediction in the world sub-model (WSM) and the other corresponding to a simulation/prediction in the linguistic sub-model (LSM). For a sentence to 'make sense', there needs to be a synchronous, stepwise prediction in both sub-models. In a different context, Altmann and Mirkovic (2009) speak of a dual, synchronous "unfolding" of sentence and real-world event representation, which fits nicely with the way I suggest we should consider the relation between the LSM and WSM:

> [...] language is not processed in isolation of the world it describes; instead, comprehension consists in realising a mapping between the unfolding sentence and the event representation corresponding to the real-world event that is being described. (Altmann & Mirkovic 2009: 602)

In this way, the WSM can be seen to provide conceptual or categorical constraints with regard to how words can be combined grammatically to create *meaningful* sentences. I suggest that liar sentences are meaningless because liar situations, on very careful reflection carried out in the WSM, cannot arise in the way that is necessary for a paradox, i.e., a synchronously unfolding prediction in the WSM and LSM fails. But still, the liar paradox retains a certain psychological pull, as it looks meaningful. This, as I will show, can be explained by an accommodation mechanism fleshed out via the PP-specific precision weighting mechanism and its multi-level processing architecture. But first I need to establish a basic assumption that is necessary for my treatment of the liar paradox, the assumption that meaning is *speaker meaning*.

4.1. An assumption: meaning as speaker meaning

I make one substantial assumption, namely, the Austinian-Wittgensteinian view of meaning as speaker-intended meaning. A defence of this view would exceed the scope of this paper, but fortunately, the view is endorsed by others as well, and I refer to the literature for further defence (e.g., Goldstein 1981, 1982, Rayo 2013, Azzouni 2013). It is by using a word or sentence in a specific context with a particular intention that the speaker endows this word or sentence with meaning. Without the (at least imaginable) presence of a (conscious, intentional) speaker, sentences are merely dead wiggles and noises. The impression that *The cat is on the mat* means something 'by itself' as a sentence without actually (or imaginatively) being uttered by a speaker with assertive intentions is merely an illusion. Azzouni (2013) explains the illusion that linguistic items mean by themselves through the phenomenon of "semantic perception". In the same way that we cannot avoid 'seeing' the functionality of a screwdriver, we cannot avoid 'perceiving' the meaning of a word:

> We don't experience the presence of a speaker's intentions -- even when we are aware of them -- as causing or influencing or determining what is said in these cases. Rather, we simply experience the expression as just meaning this or as just meaning that. (Azzouni 2013: 130)

If random wind movements arrange some leaves on the ground in the form of the word *cat*, it has a meaning only because we can imagine a person intentionally arranging those leaves to form the word *cat* to communicate something. If all meaning is speaker meaning, the test for the meaningfulness of a sentence--with careful thinking--is a speech act simulation where an agent with a certain intention is uttering the sentence. I speculate that Azzouni's "semantic perception" involves unavoidably and subconsciously projecting certain intentions onto a parrot uttering a 'meaningful' sentence.

In the view defended here, the place where a person generates and processes meaning is the brain's WSM. WSM-based meaning processing, when coordinated with the LSM, not only includes the comprehension of the literal (or semantic) meaning of words and individual sentences, but also comprises the derivation of the pragmatic meaning of a sentence, and of the understanding of whole discourses. Pragmatic meaning comprehension involves perspective-taking on the listener's side, i.e., the simulation of speech-acts and speaker intentions, from the perspective of the speaker (e.g., van Berkum et al. 2008). I also assume that the comprehension of the literal or semantic meaning of sentences requires such simulation-inferences. That might be controversial but is compatible with the Davidsonian view (e.g. 1986) which denies the existence of fixed, conventional (literal) word meanings. Such a view is gaining momentum among philosophers, psychologists, linguists and cognitive scientists (e.g., Barsalou 2009, 2011, Casasanto & Lupyan 2015, Ludlow 2014, Rice 2016) who take concepts to be highly dynamic and context-dependent entities. The view does not deny that a semantic/pragmatic distinction might be useful, but certainly questions a clear, dichotomic divide. Such a divide would seem arbitrary in the multi-level processing architecture of PP, where inferences are carried out on different levels of abstraction and contextual complexity.

The view that semantic and pragmatic processing are not separate (and sequential) cognitive processes is supported by recent empirical findings that show that meaning processing is immediate and holistic (e.g., Hagoort & van Berkum 2007, Bašnáková et al. 2014). This finding fits with the PP model, because error minimization can be seen as a holistic and simultaneous process that brings the network on all levels into an overall optimal state. Crucial for the rest of the paper is the assumption that linguistic processing in the PP model can have different 'levels of depth' (I follow Barsalou in the use of the notions 'shallow' and 'deep' here, see section 3.2). For instance, meaning typically described as pragmatically inferred is deeper than that described as literal meaning in the sense that the former integrates a larger amount of available information, namely, additional contextual and situational cues. Even the processing of what is normally described as literal (or semantic) meaning of a sentence can have different levels of depth. If conceptual knowledge is represented in the WSM in terms of constraints for concept application, then the level of depth of processing is correlated with how exhaustively all those constraints are being taken into account.

With those clarifications about the notion of 'meaning' and assumptions about how meaning is processed at hand, we can now move to the discussion of how the liar paradox arises.

4.2. Two ways of conceiving of the liar paradox

Let us call a scene or situation that leads to a liar paradox a 'liar situation'. The situation in which Mary utters the liar sentence, *I am now lying* is an example of a liar situation. Another, non-self-referential, liar situation was first described by Stephen Yablo.[5] Consider the following infinite chain of numbered sentences, each of which states that all of the sentences after it are false. As can be easily verified, no consistent assignment of truth values can be made to the sentences:

0. All following sentences (1, 2, 3, ...) are false.
1. All following sentences (2, 3, 4, ...) are false.
2. All following sentences (3, 4, 5, ...) are false.

...

I suggest that the liar paradox arises as a failure of the synchronization between two ways of generating (predicting) a liar situation. Those two ways correspond to predictions in the sub-models, the WSM and the LSM.

4.2.1. Conceiving in the LSM (linguistically focused prediction)

The first way of conceiving of the liar situation corresponds to relatively 'shallow processing' (see also Erickson & Mattson 1981, or Barton & Sanford 1993). The focus of the mind is on the LSM, not on the 'deep meanings' and implicit conceptual constraints contained in the WSM. To show the generality of the approach I will use the above examples of self-referential and non-self-referential liar situations and try to sketch what is going on inferentially when we conceive of the liar paradox in this mode.

[5] Yablo (e.g., 1985) discovered this version of the liar, the so-called *Yablo Paradox*, which shows that the problem with the liar is not self-reference!

Self-referential liar situation

We have a person, Mary, uttering a sentence (the liar sentence). Her sentence is grammatical, and it contains only meaningful words. The subject matter of what Mary utters can be easily identified. She is saying something about the truth of the utterances of an agent. The agent happens to be herself. Also, the sentence does not contain any apparent category mistakes because we can clearly say of persons, including Mary, that they say the truth or lie. The sentence appears to be meaningful. The PP model has generated the sentence with some (shallow, as we will see) involvement of the WSM. In this way, the PP model produced 'comprehension' of the sentence. The paradox arises then when we take the apparently meaningful sentence to be an assertion and start reasoning with it.

Non-self-referential liar situation: Yablo's paradox

All sentences are grammatical and contain only meaningful words. The subject matter of each sentence is easy to identify. Each sentence is expressing something about the truth or falsehood of the sentence that follows. Also, the sentences do not contain any (apparent) category mistakes because one can say about sentences that they are true or false. All sentences appear to be meaningful. The PP model has generated the sequence of sentences with some (shallow, as we will see) involvement of the WSM. In this way, the PP model produced 'comprehension' of the sequence of sentences. The paradox arises then when we proceed and try to assign consistent truth value assignments to the apparently meaningful sequence of sentences.

4.2.2. Conceiving in the WSM (semantically focused prediction)

Surprisingly, one very simple fact about liar situations does not play any role in the literature about the liar paradox. On *careful reflection* it becomes obvious that liar situations cannot actually arise in reality, at least not in the way that is necessary for a paradox to arise. When I say liar situations cannot arise in reality, I mean that we cannot conceive--on very careful reflection, taking into account all of the conceptual constraints--of situations where the sentences are being uttered with the right intentions. But this means that they are meaningless--at least under the assumption discussed above that meaning is speaker meaning. The 'on careful reflection' corresponds to the second way of conceiving of the liar situation. It is cognitive processing of the liar situation with a strong focus on the WSM, i.e., with an emphasis on 'simulating deeply' (not shallowly) an actual liar situation. Let me explain.

Self-referential liar situation

Mary makes an (apparent) assertion about her trustworthiness. Given this context, there is, however, no conceivable situation in which Mary would say, *I am lying*. If Mary were telling the truth she would say, *I am telling the truth*, and if she were lying, she would say that same sentence. Of course, the WSM and LSM can generate a scene (through a coordinated unfolding of the predictions in each sub-model) where Mary utters the liar sentence. However, in the WSM prediction Mary's intention cannot be the one required to make the liar sentence a paradoxical sentence, namely, the intention to say something true or false about her trustworthiness. As a consequence, the WSM cannot generate the liar situation synchronously with the LSM in the way required to make it a paradox. Therefore, the liar sentence is meaningless. To find the liar sentence meaningful requires being slightly sloppy with the WSM and ignoring the subtle role of the utterer's intention. How this 'sloppiness' might arise in PP terms I will discuss in a moment.

Non-self-referential liar situation (Yablo's paradox)

In the Yablo paradox, there are no agents with intentions. However, I have assumed that to make sense of those sentences, we need to be able to at least imagine a situation where intentional agents utter them with the appropriate (assertoric) intentions. We can transcribe Yablo's sequence in the following way:

Person 0: "What persons (1, 2, 3, ...) are saying is false."
Person 1: "What persons (2, 3, 4, ...) are saying is false."
Person 2: "What persons (3, 4, 5, ...) are saying is false."
...

A similar line of thought as with the self-referential situations shows that the Yablo situation cannot be generated with deep (as opposed to shallow) involvement of the WSM either. To make sense of the sequence of sentences, we need to imagine them to be uttered by agents with appropriate intentions. To say something true or false intentionally about a sentence requires knowing the sentence. For instance, if everyone speaks at the same time, they all make statements without knowing what the others are saying. So, they cannot make honest (or dishonest) assertions about the truth value of the others' sentences. It is also not possible that everyone speaks in sequence, such that they all refer only to known sentences. The reason is that in the Yablo situation, the last person should start speaking then. However, there is no last person in this infinite sequence. Therefore, 'comprehension' of the Yablo situation involves strong attention being paid by the mind to the formal aspects (i.e., a strong focus on the LSM), including the recognition that the sentences are in correspondence with a mathematical structure (an infinite omega sequence). The apparent 'comprehension' of the Yablo situation succeeds only by partially ignoring constraints from the semantic engine (the WSM). No synchronous generative unfolding with deep involvement of both sub-models is possible. Hence the sequence of sentences as a whole (and at its highest level of integration) is meaningless.

In both cases, the self-referential and non-self-referential, the liar sentences can be uttered, of course, but they cannot be asserted, where assertion implies that the utterer intends to say something true or false. The meaninglessness consists in the fact that words are put together in a way that cannot correspond to a situation generated synchronously with deeply (and not only shallowly) involving the WSM. The WSM therefore provides constraints on how words can be put together meaningfully to describe whole situations. While the liar sentence (or a sequence) as a whole is meaningless, it certainly is not completely meaningless. It is significantly more meaningful than linguistic forms like: *Dsad djdjdj hhd* or *Green go or having*. The multi-level processing of the PP model can accommodate the intuition that meaning comes in degrees.[6] The liar sentence contains meaningful words; it contains meaningful phrases; a subject matter can be clearly identified; it can be identified as a sentence with a subject and predicate, etc. Cognitive processing succeeds at those lower levels. It is only at the highest level of integration in the WSM of all those bits and pieces that meaning fails. However, this failure is enough to deprive the liar sentence of the status of being meaningful as a whole and any truth evaluation is pointless.

---

[6] Thanks to Andy Clark for this suggestion.

4.2.3. Accommodation effects: explaining the psychological pull of the liar

I will now explain in more detail how, and in what sense, the liar sentence *appears* to be meaningful. This is essential to account for the pull of the liar paradox. The pull of the liar paradox consists in the resistance to accepting that the liar sentence is meaningless. (Even I still cannot believe that the liar sentence is meaningless!)

The brain's model exhibits flexibility in interpreting other speakers of the language community. A precise application of grammar and orthography is not necessary to be able to understand others. We can perfectly well understand grammatically incorrect sentences. We can even deal with words that are used incorrectly--malapropisms (Davidson's classic paper "A nice derangement of epitaphs" (1986) makes this last point). Let me call such effects *accommodation effects*. Accommodation can be explained in the PP framework via the precision-weighting mechanism in the following way. The PP precision weighting mechanism allows us to tune up or down the influence of the error determined in the error unit corresponding to a certain representation. With this mechanism one can modulate representations (predictions) by switching off or dampening the influence of other representations.[7] Accommodation, for instance of a malapropism, occurs when the error signal for that word is ignored or dampened. In this case, top-down influence of the prediction of the right word (which is inferred from the context in which the word stands in the sentence and of the speech act itself) dominates and the prediction succeeds, despite a large error between the predicted word and the word actually uttered by the speaker.[8] *Metaphors* can also be explained by an accommodation mechanism cashed out in PP terms. Metaphors are often taken to be category mistakes, i.e., grammatically correct sentences that are made up entirely of meaningful words but are semantically anomalous. Shakespeare's famous *Juliet is the sun* is a simple and often analysed example (see, e.g., Schroeder 2004). However, the PP model can explain metaphors via accommodation. Metaphors can be 'made' meaningful by flexibly creating an ad-hoc concept to 'force a

---

[7] See Michel (2020) for a proposal of a detailed mechanism for context-sensitive modulation of concept features with PP precision weighting.

[8] Another example of an accommodation process, quite close to malapropisms, is *Moses sentences* (see also Erickson & Mattson 1981, or Barton & Sanford 1993): people tend to answer to the question *How many animals of each kind did Moses take on the Ark?* with *Two*. They overlook that it should say *Noah*.

fit'. Features of the concept SUN are suppressed via the precision weighting mechanism and the ad-hoc concept *SUN is created, which can represent a more inclusive and abstract concept composed of features like 'something central' or 'something that generates positive feeling'. Not all category mistakes can be accommodated as metaphors. For instance, *The relativity theory listens to an ill breakfast* seems to be nonsense that cannot be accommodated, i.e., 'forced' to be meaningful.[9]

Accommodation can be applied to the liar sentence as well. That the liar sentence appears to be meaningful might be explained in the following way. On the one hand, the sentence is fully grammatical; the agent appreciates this by focusing attention on the LSM, in which the sentence can be easily generated. For the sentence to also be meaningful to the agent requires success through a synchronous, parallel unfolding of the prediction in the WSM while the stepwise prediction in the LSM is going on. While the synchronous unfolding cannot succeed on 'careful, constrained conceiving' (i.e., deep as opposed to shallow processing) in the WSM, it can succeed if we raise the error thresholds for the prediction, in other words, if we simulate the liar scene with a coarser grain, ignoring certain features of the liar situation. This effect might lead to abstracting away the speaker's intentions. In fact, in the overwhelming majority of cases we can ignore the speaker's intentions and consider sentences as having meaning on their own. Our bias towards perceiving meanings as properties of sentences (see the idea of *semantic perception* described by Azzouni, mentioned before) might be an evolutionary adaptation. In most cases it works. However, on some occasions, like the liar situations, it leads to trouble.

One might object to the fact that we have left unexplained the exact sort of conceptual constraint that makes the sentences in liar situations meaningless. But this is only a pressing concern for those who think that we must be able to formalise and make explicit the conceptual constraints in the PP models (which I suggested are encoded in the WSM sub-model). Although liar situations are quite simple situations, they involve a complex web of common-sense concepts related to language and concepts like truth and falsehood, as well as folk-psychological knowledge of intentions, etc. It is unclear how those could be formalised explicitly in the way that formalists demand, such that

---

[9] Carston (2002) also suggested that some metaphors can be explained by the creation of ad-hoc concepts where content is narrowed or broadened, until one can 'make sense' of the metaphorical sentence.

we can filter out all possible liar situations. In section 5.2. I will delve further into this issue.

Let me stress that the treatment of the liar paradox requires the specific PP apparatus. Specifically, the mechanism of precision weighting is central to the explanation. The ultimate purpose of the overall PP model is "to get a grip on the world" (Clark 2016: 202), the central tenet of PP. The LSM could be seen as a model of the WSM (and hence as a second-order model of the world) that can be made explicit linguistically. It enhances the grip-getting capacity and efficiency of the WSM in many ways, e.g., by allowing the sharing of predictions via public language or the critical evaluation of predictions that are publicly available (see Clark 1998 for various other ways applicable also to PP). But the LSM can also lead us astray, as the liar paradox shows, if we reason with the LSM with only shallow involvement of the WSM, i.e., by not carefully honouring all of the conceptual constraints within it.

A further conclusion is that the multi-level processing of the PP framework offers resources to deal with partial meaningfulness, which we have encountered in the liar paradox. An application of those ideas in non-linguistic contexts also seems possible. For instance, take Escher's famous vexing visual paradoxes. Escher pictures present 'impossible' three-dimensional objects (e.g., Escher 1971, Penrose & Penrose 1958). A lot of local and partial patterns can be recognised as meaningful without any problem; however, no stable perception of the whole three-dimensional object is possible. Unfortunately, detailed discussion of this phenomenon needs to be carried out elsewhere. But a treatment similar to the liar paradox seems possible along those lines.

5. An objection: the *Principle of Unrestricted Compositionality*

I will focus on what I consider the most serious possible objection to the view that liar sentences are meaningless. This objection is related to what I suggest calling the *Principle of Unrestricted Compositionality* (PUC). The PUC holds that all grammatical sentences with meaningful words are meaningful and truth-evaluable. (Specifically, category mistakes come out simply as false under the PUC.) As liar sentences are grammatical and have only meaningful words, then if the PUC is correct, my treatment of the liar paradox must be flawed, as it relies on liar sentences being meaningless.

The PUC is motivated by the *productivity* and *systematicity* of language. The *productivity* ensures that we can generate infinite novel combinations of concepts. If we can entertain the thought a is G, then we can also entertain that b is G, c is G, etc. If we can entertain that a is F then we can entertain that a is G for every concept of G we have. The *systematicity* allows us to understand that Peter kisses Mary when we understand that Mary kisses Peter. The PUC goes back to Evans' (1982) *generality constraint*:

> If a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception. (Evans 1982: 104, as cited in Camp 2004)

Evans stresses that for an agent to competently master a particular concept, she needs to be able to combine it arbitrarily with any other concept. But the PUC has not been the mainstream view. Other writers (e.g., Strawson 1959, Peacocke 1992) prefer to impose categorical constraints on concept combinations to avoid 'category mistakes' like *Colorless green dreams sleep furiously* (Chomsky 2002: 15). The strong intuition is that those are not meaningful sentences; one cannot grasp under which conditions those sentences would be true. The inability to understand what is being said in cases of absurd concept combinations should not undermine the conceptual competence of the agent. But Camp (2004) and Magidor (2013) have recently put forward a whole battery of arguments defending the PUC. According to them, even category mistakes like *Caesar is a prime number* are meaningful and can be truth-evaluated. Let me briefly discuss three of their arguments.

5.1. Argument from (material) inferential roles of category mistakes

Camp (2004: 212) suggests that category mistakes have substantial *inferential roles* and, therefore, are meaningful and can be truth-evaluated. She argues for this thesis in the following way. From the supposition that Caesar is a prime number you can draw the material inference that 'Caesar is not an efficacious emperor'. This inference, according to Camp, is material, not formal, because it requires using the meanings of the terms involved. If the inference were merely formal, we could replace subject and objects with constants and ignore meanings and draw the inference in a purely mechanical way. The conclusion clearly makes sense; therefore, if one can infer meaningful things from category mistakes, then category mistakes cannot be meaningless.

However, this argument begs the question. One can only make the presupposition ‘Caesar is a prime number’ if that sentence is meaningful. But that is what Camp wants to show. Now, maybe she might respond that I beg the question by taking the sentences as meaningless and for that reason reject the idea that it can be taken for a supposition for the sake of argument. So, I first have to show on independent grounds that the sentence is meaningless. However, even if we could presuppose that Caesar were a prime number, what Camp calls a substantial material inference is nothing more than what could be called inferential luck because the conclusion is meaningful for the wrong reason; also, the inference turns out to be formal and not material (i.e., meaning involving) after all. To see this, let us spell out the inference in more detail (this time assuming, for argument’s sake, that Caesar is a prime number):

> Assumption: Caesar is a prime number.
> P1: If something is a prime number, then it is abstract.
> P2: Abstract things are not efficacious.
> P3: Prime numbers are not efficacious (by virtue of being abstract)
> -> Caesar is not efficacious (because of his abstractness in virtue of being a prime number).

We have a meaningful conclusion for the wrong reason. We have been able to conclude something that obviously makes sense because non-efficaciousness can be meaningfully predicated on both persons and abstract things. The simple recipe for further examples is as follows: assume that X is a Y (where “X is a Y” expresses a category mistake). Take some predicate P that can be meaningfully predicated on Xs and Ys. Then the inference runs:

> Assume X is Y.
> Y is P.
> Hence X is P.

As this scheme indicates, we do not need the meaning of X, Y, and P, merely the condition that P can be predicated on both X and Y. So, the inference is rather non-material and formal.

5.2. Argument from failure to formalise categorical restrictions

Magidor (2013:48) argues that the best attempt at formalising categorical restrictions in natural language (namely, via *Montague grammar*) has failed. No other formalisation is forthcoming; hence there are no good reasons to think that categorical constraints can be formalised.

However, Magidor has not shown that categorical restrictions cannot be formalised, just that they cannot be formalised in some simple and elegant way as *type theory* aims to do (type theory builds on a few, two or so, “types”--like “individuals” and “truth values” from which further types are built by combinations) (Magidor 2013: 48–56). Magidor still leaves room for very gerrymandered restrictions or having one specific type for almost each concept (2013: 5–-52). My response is that I would expect from the human conceptual apparatus that its categorical restrictions would turn out to be quite messy. An elegantly formal model, inspired by static axiomatic systems or traditional semantic toy models, also seems inadequate. The human conceptual apparatus is highly flexible and dynamic and conceptual content associated by tokening a concept seems highly context-dependent (e.g., Barsalou 2009, 2011, Casasanto & Lupyan 2015, Ludlow 2014, Rice 2016). Furthermore, as I have pointed out in the introduction, there is strong evidence for concepts involving modality-specific representations, i.e., non-language-like representations. So, it is not clear how such a neat formalisation could be carried out in a language-like format at all. Probably the representational system that can be formalised is some approximation, but the brain certainly does not implement a formal calculus exactly.

While I do not think that the actual conceptual apparatus can be neatly formalised for descriptive purposes, it is possible that our generative model in the brain contains some very high-level ‘hyper-prior’, i.e., some sub-personal, internal ‘norm’, that makes us unconsciously ‘desire’ (especially in scientific contexts) to have our mental representations and concepts formalised. The fact that we so persistently undertake--in scientific contexts--axiomatisations and formalisation might be a symptom of such an unconscious bias.

5.3. Argument from scientific progress

Camp (2004:230) further argues that scientific progress sometimes consists in formulating hypotheses that look senseless and are only minimally understood, for example, that light is a wave and a particle at the same time.

However, I do not think that this is an argument for unconstrained compositionality. There are at least two responses that allow for avoiding the acceptance of the PUC. Firstly, what Camp's example shows is that there can be *concept change* and that accommodation mechanisms are at work (like in the case of metaphors). The pre-theoretic concept of light is not dualistic, while the revised, scientific concept is. Of course, the first time one spells out the idea that light is both a wave and a particle it sounds like nonsense (which it is--according to the old concept of light), and unless the idea is spelled out more specifically, leading to concept change, it remains nonsense. The fact that a meaningless sentence can trigger a concept revision does not imply that it was meaningful before the conceptual change happened. Concept change might be considered what I call an accommodation process over some larger time scale and with more permanent effects on the functional webs by which concepts are implemented. The connections of the (pre-change) concept of light to other concepts might be adjusted in such a way that taking light to be both a particle and wave 'makes sense' again. A second response (thanks to Andy Clark, in a personal conversation) is to insist that the dual concept of light is really a sort of disjunctive concept, where we have two partial models, one with light as a particle and the other with light as a wave. Science has not really integrated those partial models into one unified concept.

A last comment about the PUC as a mark of the conceptual competence of speakers is in order. It seems to me that avoiding meaningless concept combinations (and finding them 'meaningless') is a sign of conceptual competence, not the incompetence of the speaker. The skill of correctly combining concepts demonstrates that she knows the categorical restrictions. The meaning of concepts is implicitly encoded in the categorical restrictions. Someone systematically combining words in nonsensical cross-categorical ways would probably not be judged competent in the language. Also, the PUC, rather than a requirement for *conceptual* competence, is a requirement of *grammatical* competence, because it is knowledge of the grammatical structure that underlies the possibility of unconstrained combinability. Semantic knowledge consists in the skill of being able to select those combinations that make sense.

I conclude that Camp's and Magidor's arguments for the PUC, and hence against the meaningless view of category mistakes, are not decisive. Therefore, the meaninglessness solution to the liar paradox, for which I have tried to provide a supporting cognitive account, is not undermined.

6. Conclusion

Firstly, I motivated a cognitive approach to understanding and resolving the liar paradox in opposition to the popular formalistic approaches. Building on a specific cognitive computational architecture for the mind, based on the Predictive Processing framework, I argued that the liar sentence is meaningless. It is meaningless because the situation that the sentence evokes cannot be conceived of in the 'world model' of the brain (in a way that is necessary for a paradox). Still, the liar looks meaningful and it is difficult to escape its grip. This pull I explained via an accommodation process, fleshed out with a PP-specific mechanism of precision weighting of prediction errors. If the liar sentence is indeed meaningless (at the highest level of integration), the Principle of Unrestricted Compositionality must be wrong. I tried to show that some recent arguments for that principle are not decisive and do not threaten the meaninglessness solution.

Acknowledgements

I would like to thank Andy Clark, Mark Sprevak, Brian Rabern and two anonymous reviewers for their very helpful comments and suggestions.

References

Allopenna, Paul D., James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*. 38(4). 419–439.

Altmann, Gerry T. M. & Jelena Mirković. 2009. Incrementality and prediction in human sentence processing. *Cognitive Science*. 33(4). 583–609.

Azzouni, Jody. 2013. *Semantic perception: how the illusion of a common language arises and persists*. Oxford: Oxford University Press.

Barsalou, Lawrence W. 2009. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 364(1521). 1281–1289.

Barsalou, Lawrence W. 2011. Integrating Bayesian analysis and mechanistic theories in grounded cognition. *Behavioral and Brain Sciences*. 34(04). 191–192.

Barsalou, Lawrence W., Ava Santos, W. Kyle Simmons & Christine D Wilson. 2008. Language and simulation in conceptual processing. In Manuel de Vega, Arthur Glenberg & Arthur Graesser (eds.), *Symbols and embodiment: Debates on meaning and cognition*, 245–283. Oxford: Oxford University Press.

Barton, Stephen B. & Anthony J. Sanford. 1993. A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*. 21(4). 477–487.

Barwise, Jon & John Etchemendy. 1987. *The liar: an essay on truth and circularity*. New York: Oxford University Press.

Bašnáková, Jana, Kirsten Weber, Karl M. Petersson, Jos van Berkum & Peter Hagoort. 2014. Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*. 24(10). 2572–2578.

Beall, Jc, Michael Glanzberg & David Ripley. 2019. "Liar Paradox", *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/liar-paradox/>. (12 March, 2020).

Berkum, Jos J. van, Danielle Van den Brink, Cathelijne M. Tesink, Miriam Kos & Peter Hagoort. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience*. 20(4). 580–591.

Camp, Elisabeth. 2004. The generality constraint and categorial restrictions. *The Philosophical Quarterly*, 54(215). 209–231.

Carston, Robin. 2002. *Thoughts and Utterances*. Oxford: Blackwell.

Casasanto, Daniel & Gary Lupyan. 2015. All concepts are ad hoc concepts. In Eric Margolis & Stephen Laurence (eds.), *The conceptual mind: New directions in the study of concepts*, 543–566. Cambridge, MA: MIT Press.

Chomsky, Noam. 2002. *Syntactic structures*, 2nd edn. Berlin & New York: Mouton de Gruyter.

Clark, Andy. 1998. Magic words: How language augments human computation. In Peter Carruthers & Jill Boucher (eds.), *Language and thought: Interdisciplinary themes,* 162–183. Cambridge: Cambridge University Press.

Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. 36(03). 181–204.

Clark, Andy. 2016. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford, New York: Oxford University Press.

Dam, Wessel O. van, Margriet van Dijk, Harold Bekkering & Shirley-Ann Rueschemeyer. 2012. Flexibility in embodied lexical-semantic representations. *Human Brain Mapping*. 33(10). 2322–2333.

Dam, Wessel O. van, Eelco V. van Dongen, Harold Bekkering & Shirley-Ann Rueschemeyer. 2012. Context-dependent changes in functional connectivity of auditory cortices during the perception of object words. *Journal of Cognitive Neuroscience*. 24(10). 2108–2119.

Davidson, Donald. 1986. A nice derangement of epitaphs. In Ernest Lepore (ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson.* 433–446. Oxford: Blackwell.

Dell, Gary S. & Franklin Chang. 2013. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 369(1634).

Dutilh Novaes, Catarina. 2012. *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge, New York: Cambridge University Press.

Elqayam, Shira. 2006. The collapse illusion effect: A semantic-pragmatic illusion of truth and paradox. *Thinking & Reasoning*. 12(2). 144–180.

Erickson, Thomas D. & Mark E. Mattson. 1981. From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*. 20(5), 540–551.

Escher, Maurits C. 1971. *The graphic work of M. C. Escher*. New York: Ballantine.

Evans, Gareth & John H. McDowell. 1982. *The varieties of reference*. Oxford: Clarendon Press.

Fodor, Jerry A. 1979. *The Language of thought*. Cambridge, MA: Harvard University. Press.

Fodor, Jerry A. & Zenon W. Pylyshyn. 2015. *Minds without meanings: an essay on the content of concepts*. Cambridge, MA: The MIT Press.

Fraassen, Bas van. 1970. Truth and Paradoxical Consequences. In Robert Martin (ed.), *The paradox of the liar,* 13–23. New Haven: Yale University Press.

Friston, Karl. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*. 11(2). 127–138.

Gagnepain, Pierre, Richard N. Henson & Matthew H. Davis. 2012. Temporal Predictive Codes for Spoken Words in Auditory Cortex. *Current Biolog*y. 22(7). 615–621.

Gerstenberg, Tobias & Joshua B. Tenenbaum. 2017. Intuitive theories. In Michael R. Waldmann (ed.), *Oxford Handbook of Causal Reasoning*, 515–548. New York, NY: Oxford University Press.

Goldstein, Laurence. 1981. Categories of linguistic aspects and Grelling's paradox. *Linguistics and Philosophy*. 4(3). 405–421.

Goldstein, Laurence. 1982. Linguistic aspects, meaninglessness and paradox: A rejoinder to John David Stone. *Linguistics and Philosophy*. 4(4). 579–592.

Hagoort, Peter & Jos van Berkum. 2007. Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 362(1481), 801–811.

Hoenig, Klaus, Eun-Jin Sim, Victor Bochev, Bärbel Herrnberger & Markus Kiefer. 2008. Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*. 20(10). 1799–1814.

Hohwy, Jacob. 2013. *The predictive mind*. Oxford, UK, New York, NY: Oxford University Press.

Kearns, John T. 2007. An illocutionary logical explanation of the liar paradox. *History and Philosophy of Logic.* 28(1). 31–66.

Kripke, Saul. 1975. Outline of a theory of truth. *The Journal of Philosophy*. 72(19). 690–716.

Kuperberg, Gina R. & Tim F. Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*. 31(1). 32–59.

Kutas, Marta, Katherine A. DeLong & Nathaniel J. Smith. 2011. A look around at what lies ahead: Prediction and predictability in language processing. In Moshe Bar (ed.), *Predictions in the brain: Using our past to generate a future*, 190–207. Oxford: Oxford University Press.

Lakoff, George & Rafael E. Núñez. 2000. *Where mathematics comes from: how the embodied mind brings mathematics into being*, 1st edn. New York, NY: Basic Books.

Langacker, Ronald W. 2008. *Cognitive grammar: a basic introduction*. Oxford & New York: Oxford University Press.

Leitgeb, Hannes. 2007. What theories of truth should be like (but cannot be). *Philosophy Compass*. 2(2). 276–290.

Ludlow, Peter. 2014. *Living words: meaning underdetermination and the dynamic lexicon*, 1st edn. Oxford: Oxford University Press.

Lupyan, Gary & Andy Clark. 2015. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science.* 24(4). 279–284.

Magidor, Ofra. 2013. *Category mistakes*. Oxford: Oxford University Press.

Martinich, Aloysius. P. 1983. A Pragmatic solution to the Liar Paradox. *Philosophical Studies*. 43(1). 63–67.

Michel, Christian. (2020). Concept contextualism through the lens of predictive processing. *Philosophical Psychology*. 33(4). 624-647.

Paivio, Allan. 1990. *Mental representations: A dual coding approach.* Cary: Oxford University Press. Retrieved from http://public.eblib.com/choice/publicfullrecord.aspx?p=3052439. (12 March, 2020).

Peacocke, Christopher. 1992. *A study of concepts*. Cambridge, MA: MIT Press.

Penrose, Lionel S. & Roger Penrose. 1958. Impossible objects: A special type of visual illusion. *British Journal of Psychology*. 49(1). 31–33.

Pickering, Martin J. & Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*. 36(04). 329–347.

Priest, Graham. 1984. Logic of paradox revisited. *Journal of Philosophical Logic*. 13(2). 153–179.

Priest, Graham. 2006. *In contradiction: A study of the transconsistent*. Oxford: Clarendon Press.

Rayo, Agustin. 2013. A plea for semantic localism. *Noûs*. 47(4). 647–679.

Rice, Collin. 2016. Concepts as Pluralistic Hybrids. *Philosophy and Phenomenological Research*. 92(3). 597–619.

Ripley, David. 2016. Experimental philosophical logic. In Justin Sytsma (ed.), *A companion to experimental philosophy*, 523-534. Wesley Buckwalter.

Rips, Lance J. 1989. The Psychology of Knights and Knaves. *Cognition.* 31(2). 85–116.

Schroeder, Severin. 2004. Why Juliet is the Sun. In Mark Siebel and Mark Textor (eds.), *Semantic and ontology,* 63–100. Frankfurt a.M: Ontos Verlag.

Simmons, W. Kyle, Stephan B. Hamann, Carla L. Harenski, Xiaoping P. Hu & Lawrence W. Barsalou. 2008. fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*. 102(1–3). 106–119.

Strawson, Peter F. 1959. *Individuals: An essay in descriptive metaphysics*. London: Methuen & Co. Ltd.

Tversky, Amos, & Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*. 185(4157). 1124–1131.

Williamson, Timothy. 1996. *Vagueness*. New York: Routledge.

Yablo, Steven. 1985. Truth and reflection, *Journal of Philosophical Logic.*14. 297–349.