Belief, Truth, and Ways of Believing

Chapter 7 in Modes of Truth, C. Nicolai and J. Stern (eds), Routledge

Johannes Stern
Department of Philosophy
University of Bristol
johannes.stern@bristol.ac.uk

Abstract

The notions of belief and truth frequently interact in philosophical discourse but, surprisingly, a coherent semantics for such discourse is still wanting. Indeed, a number of puzzles stand in way of a satisfactory semantic account of the notion of truth in doxastic contexts. In this paper we discuss these puzzles and develop a more satisfactory semantic account that combines ideas from contextualist theories of attitude reports and Awareness semantics for non-idealized belief.

1 Introduction

Philosophy without truth, knowledge and belief would be a fairly boring discipline—there would only be the good and the beautiful left to discuss. Fortunately, philosophy is exciting and truth, knowledge, and belief are notions at the center of the discipline responsible for many important philosophical questions and puzzles. The three notions are intimately connected and, as a consequence, so will be the philosophical questions and puzzles of the respective notions. For example, knowledge guarantees truth, i.e., it is factive and, indeed, it is arguably at least in parts this characteristic that distinguishes knowledge from mere belief. Whether this means that knowledge can be defined on the basis of knowledge, truth and, possibly, some further condition has been the question shaping much of the recent debate in epistemology. With this observation in mind one would expect that most formal philosophizing is conducted in a formal framework in which truth, knowledge and belief are treated simultaneously, so the formal and philosophical views regarding the connection of the different notions can be tested for their consequences. Surprisingly, no satisfactory such framework has—to our knowledge—been developed to date. Of course, starting with Hintikka (1962) there has been a lot of work on formal semantics and logics of knowledge and belief but unfortunately very little work on how to construct an adequate theory of truth in these contexts.[1] The aim of this paper is to take first steps in developing a satisfactory formal framework in which contemporary debates

[1] Caie (2012); Jerzak (2019) and, arguably, Halbach and Welch (2009); Campbell-Moore (2015); Stern (2016) are notable exceptions to this claim. Yet, the semantics presented by all these authors produce the type of unintended consequence discussed in Section 2.

in epistemology can be aptly represented. To this end, we focus on the notion of truth in belief contexts—although a number of observations would also apply to the interaction of the notions of truth and knowledge—and start by examining a major hurdle or puzzle in way of a satisfactory semantics for truth in doxastic contexts. We then analyze the philosophical underpinnings of the puzzle and develop a semantics for the notion of truth in doxastic contexts, which is based on our analysis. We discuss some of its consequences and, before concluding, point to some limitations of the semantics and outline some alternative strategies for developing adequate semantics for truth in doxastic contexts.

2 Semantics for Truth and Belief: Overgeneration

As mentioned, Hintikka's seminal *Knowledge and Belief* (Hintikka, 1962) lay the foundation for formal philosophizing about knowledge and belief. Hintikka proposed a formal interpretation of belief and, respectively, knowledge within the framework of possible world semantics (henceforth PWS). According to Hintikka an agent believes (knows) that φ if and only if 'φ' is true at all of the agent's doxastic (epistemic) alternatives—worlds that are accessible via the doxastic (epistemic) accessibility relation—where the truth predicate is understood in a metalinguistic, that is, model-theoretic sense. Most subsequent work on formal semantics for belief has followed Hintikka's footsteps in analyzing belief in some form of possible world semantics broadly conceived, that is, as some form of quantifier over worlds, states, or situations.[2] It seems only reasonable then to take the possible world analysis as a starting point for a combined formal framework for truth and belief. What we are after is a framework in which the notions of truth and belief figure in the object-language, that is, we want to formulate claims such as

(1) Not everything Boris believes is true.

As a consequence, standard PWS for epistemic notions will not be sufficient because, as mentioned, the truth predicate at play in the semantics is the metalinguistic one.[3]

If an object-linguistic truth predicate is introduced to the framework of PWS, its semantic interpretation needs to be specified, that is, the interpretation of the truth predicate at every possible world has to be determined. To this end, it is not sufficient to determine the interpretation of the object-linguistic truth predicate at a given world by fiat. Rather, if, following the outlines of a commonly accepted view on truth and paradox, semantic states of affairs supervene on non-semantic states of affairs (cf., e.g., Tarski, 1944; Kripke, 1975; Yablo, 1982; Leitgeb, 2005), the interpretation of the truth predicate at a given world should arguably depend on the interpretation of the non-semantic expressions at that world: a sentence φ will be in the interpretation of the truth predicate at a world only if the possible world models satisfies φ

[2]This is not supposed to be a controversial statement: of course, the work by theorists working with structured propositions (see Section 6) or within certain forms of truth-maker semantics may not subscribe to such an analysis. But as far as developed formal semantics go PWS is, by far, still the dominating approach.

[3]Arguably, to formulate (1) one would need to formalize belief as a predicate rather than a sentential operator as customary in PWS. We shall not discuss this issue but assume that a belief predicate can be retrieved in the language via some sort of "Kripke-reduction" (cf. Halbach and Welch, 2009; Stern, 2016, Ch. 4).

at a world w. If this idea is taken seriously, then an interpretation of the truth predicate f is adequate relative to a possible world model M and world w only if, where t_φ is a name of φ,

(TSW) $$M, w \vDash^f \mathrm{T}t_\varphi \Longleftrightarrow M, w \vDash^f \varphi.$$[4]

Fortunately, finding adequate interpretations of the truth predicate in PWS does not pose a major technical obstacle: one can simply relativize one's favorite theory of truth to the possible world framework and simultaneously construct the interpretations of the truth predicate relative to every possible world of the modal frame (cf., e.g., Kripke, 1975; Asher and Kamp, 1989; Gupta and Belnap, 1993; Halbach and Welch, 2009; Stern, 2014a,b, 2016).

The foregoing suggests that a semantics for truth in doxastic contexts can be obtained by supplementing standard doxastic PWS by an interpretation of the object-linguistic truth predicate relative to each possible world following well-rehearsed strategies discussed in the relevant literature. Unfortunately, it turns out that things are not quite as simple as that: while combining possible world semantics with standard truth-theoretic tools yields a powerful semantics for truth in belief contexts, the semantics turns out to be too powerful and to validate principles and inferences that ought not to be taken for granted. In particular, (TSW) implies that in every belief model M and world w whatever an agent believes at w, they also believe to be true and vice versa. Let's call this the Original Sin (OS) of PWS:

(OS) $$M, w \vDash^f \mathrm{B}\varphi \Longleftrightarrow M, w \vDash^f \mathrm{BT}t_\varphi.$$

(OS) will hold independently of whether we consider worlds, states or situations, as long as φ and $\mathrm{T}t_\varphi$ receive the same semantic value at these points of evaluation, that is, if (TSW) holds at every point of evaluation and the believe operator B is conceived of as a quantifier ranging over points of evaluation.[5] Notice that abandoning (TSW) ought not to be taken lightly, since, at least prima facie, this would undermine the idea that semantic states of affairs ought to supervene on non-semantic states-of-affairs. In sum, (OS) is a consequence of the two fundamental assumptions underlying PWS for the belief operator and the semantic interpretation of the truth predicate respectively.

Let us now reflect on why we should be reluctant in accepting (OS), that is, why believing and believing-true ought to be semantically differentiated. To this end, we shall present a number of cases, which, at least at the outset present counterexamples to (OS). One such case is based on the idea that the truth predicate may not be part of an agent's conceptual resources. Meet Xaver:

Xaver believes Bavaria is beautiful. But because his conceptual resources lack the truth predicate Xaver simply cannot form the belief that 'Bavaria is beautiful' is true.

[4] We do not assume that $\vDash$ is a classical satisfaction relation, that is, (TSW) is not necessarily equivalent to

(TS) $$M, w \vDash^f \mathrm{T}t_\varphi \leftrightarrow \varphi.$$

Indeed in the semantics for belief and type-free truth we shall construct $\vDash$ will be a non-classical satisfaction relation according to which (TS) and (TSW) are not equivalent.

[5] In particular (OS) holds in the semantics for belief and truth proposed by Caie (2012) and Jerzak (2019).

It seems hard to deny that it is impossible for Xaver to form an attitude towards that 'Bavaria is beautiful' is true, however, it is another question altogether whether Xaver's particular disposition amounts to a compelling counterexample to (OS). First, we may simply stipulate PWS for truth in doxastic contexts to be concerned with a theory of the doxastic attitudes of agents that have the necessary conceptual resources, i.e., conceptual resources that comprise the truth predicate. Perhaps, one might think that this condition is overly demanding or restrictive, i.e., even rational agents should not be expected to have a truth predicate at their disposition. But the aim of the semantics is not to give a general theory of attitude reports.[6] Rather the aim is to provide a semantics for truth in doxastic contexts and from this perspective it seems perfectly acceptable to focus on a semantics for agents with the necessary conceptual resources. After all, a similar kind of argument could be applied against the plausibility of any general inference involving higher-order beliefs, e.g., introspection principles—an agent may simply lack the conceptual resources to form an higher-order belief: we would be forced to conclude that the wealth of research on the plausibility of such principles is an idle exercise.

Second, even if Xaver's conceptual resources were not to include a truth predicate, this does not imply that we cannot introduce the truth predicate to the language we employ for theorizing about, or reporting, the agent's attitude. For example, meet Anne:

Anne believes Euclidean geometry to be incorrect and by modus tollens infers that one of the axioms must be incorrect without settling on one specific axiom (she may not even know all the axioms).

In this case it seems—or at least a disquotationalist would argue—that Anne's belief is correctly reported by

(2) Anne believes that not all axioms of Euclidean geometry are true;

independently of whether Anne's conceptual resources comprise the truth predicate. More importantly, at first glance it seems as if we require the truth predicate in our language to describe Anne's belief correctly. Admittedly, the view comes with important theoretical costs, namely, that the truth predicate is transparent even in highly opaque contexts but the point still stands: the absence of the truth predicate from an agent's conceptual resources is not sufficient to argue against (OS).

In sum, we take it that the charge against (OS) based on the idea that an agent's conceptual resources may lack the truth predicate to be unconvincing and will dismiss it for the purpose of our paper. But there is more damning evidence against (OS). In particular, there are more convincing cases to the effect that an agent can believe something without believing it true. Meet Clara:

Clara believes that Clark Kent is strong. But she would never express her belief in this way because she only believes that 'Superman is strong' is true. She does not believe that 'Clark Kent is strong' is true.

[6] In contrast to the Quinean analysis of attitude reports (Quine, 1956), in studying a semantics for truth in doxastic contexts there is no presupposition that all attitudinal relations implicitly appeal to the truth predicate, e.g. we do not assume the 'believes that φ' ought to be always reconstructed as 'believes-true 'φ' '.

Clara's beliefs are in plain contradiction with (OS): she believes something without believing it true. Moreover, an argument to the effect that despite appearances Clara does believe that 'Clark Kent is strong' is true and that our intuitions contradicting this assessment are down to pragmatic effects rather than a semantic distinction would seem hardly convincing in this case: at least prima facie by reporting that Clara believes that 'Clark Kent is strong' is true, we assert that Clara believes something true relative to a particular syntactic representation. The syntactic representation at stake is made explicit in the belief report and should therefore be part of the semantic content of the belief report.

Admittedly, in reporting Clara's belief we have assumed that the belief relation is merely a relation between an agent and semantic content where names are conceived as rigid designators, i.e., the syntactic or cognitive representations of the belief are not relevant for the semantic evaluation of the belief report. On alternative accounts of attitude reports, it would be incorrect to say that Clara believes that Clark Kent is strong. We take it that by constructing a semantics for truth in belief contexts, one should ideally remain neutral with respect to the particular theory of attitude reports assumed and, hence, not dismiss counterexamples to (OS) because they depend on a particular—rather popular—account of belief reports. Moreover, in general arguments against (OS) do not rely on a particular theory of attitude reports. Meet Max:

Max believes that Goldbach's conjecture is true. His friend Philip told him so and Philip is a mathematical genius. Max has absolute faith in Philip and believes him even though he has no idea what Goldbach's conjecture asserts. In fact, he does not believe that every even number > 2 is the sum of two prime numbers.

Max believes Goldbach's conjecture true without believing it. It seems undeniable that Max has not formed an attitude towards Goldbach's conjecture; he does not believe it. It also seems clear that Max believes Goldbach's conjecture is true. Perhaps one might be tempted to argue that one can only believe that Goldbach's conjecture is true if one is aware of what Goldbach's conjecture asserts. But this imposes too strict and indeed incorrect conditions on believing. We frequently believe claims, theories, etc. true without being fully aware what they assert. Moreover, we often form such beliefs simply due to (hopefully) expert testimony. In sum, we think that Max's beliefs are a clear counterexample against (OS) and that, more generally, the evidence against the semantic equivalence of believing and believing-true is damning: believing and believing-true need to be semantically differentiated.

Having corroborated the claim that the combination of possible world semantics for belief and basic desiderata regarding the interpretation of the truth predicate, when combined, yield unintended results for truth in belief contexts, the question arises whether the unintended results of the semantics, i.e. (OS), are merely a case of a formal semantics having unintended consequences or whether these results point to a deeper, philosophical problem pertaining to the notion of truth in belief contexts. In the latter case, we may yield invaluable insights for developing an adequate semantics by addressing the philosophical problem. Indeed, it turns our that the purported semantic equivalence of believing and believing-true is rooted in a philosophical puzzle about belief: if a disquotational view of truth à la Field (1994) is assumed, then the semantic equivalence of believing and believing-true is but another Fregean puzzle about belief.

3 Believing, Believing-true and a Puzzle about Belief

Traditionally, Fregean puzzles about belief employed the idea that if two names refer to the same object, they should be intersubstitutable salva veritate. But it is well known that the substitution of coreferential terms in belief contexts leads to counterintuitive consequences—indeed it were these counterintuitive consequences that led Frege (1892) to conclude that the referent of a name in oblique contexts such as belief, was not the actual referent but the sense associated with the name—and one might therefore be wary of appealing to the substitution of coreferential terms when reasoning about belief contexts. Kripke (1979) argued that the appeal to the intersubstitutivity of coreferential terms was inessential in formulating Frege-style belief puzzles. Rather Kripke based the formulation of such puzzles on two so-called disquotational principles:[7]

(DQ) If an agent A sincerely, reflectively, and competently accepts a sentence s (under circumstances properly related to a context c), then A believes, at the time of c, what s expresses in c.

(CDQ) If an agent A sincerely, reflectively, and competently denies or withholds acceptance from a sentence s (in a context c), then A does not believe, at the time of c, what s expresses in c.

At least, if agents are competent speakers of the language at stake, (DQ) and (CDQ) are *prima facie* plausible assumptions linking the acceptance of a sentence by an agent to the agent's belief in the semantic content expressed by the sentence. But if (DQ) and (CDQ) are granted, this raises two puzzles about Clara's beliefs: since, on the one hand, Clara will accept the sentence 'Superman is strong', we can infer by (DQ) that

(3) Clara believes that Clark Kent is strong.

On the other hand, since Clara will withhold acceptance to 'Clark Kent is strong', we can infer

(4) Clara does not believe that Clark Kent is strong.

by (CDQ). Moreover, (DQ) does not only imply (3) but also

(5) Clara believes that Clark Kent is not strong.

since Clara would arguably accept the sentence 'Clark Kent is not strong'. We are left with a dilemma, that is, a Fregean puzzle about belief: not only does Clara hold, in virtue of (3) and (5), mutually incompatible beliefs but we also face the question whether Clara believes that Clark Kent is strong, as suggested by (3), or not, as claimed by (4)?

[7]Here, we employ the slightly more explicit formulation given in, e.g., Nelson (2019). Kripke (1979) originally formulated the disquotational principles using 'assents to' instead of 'accepts'. Kripke also lists a number of qualifications that are intended to rule out unusual or atypical circumstances that would interfere with the agent assenting or expressing dissent with a sentence s. We implicitly adopt these qualifications.

However, the disquotational principles (DQ) and (CDQ) do not only generate Frege-style puzzles about belief, they also immediately imply that believing and believing-true are semantically equivalent, if a disquotational view of the truth predicate along the lines of Field (1994) is assumed. On such a disquotational perspective the sentence/utterance φ and the sentence/utterance $\mathrm{T}t_\varphi$ are not only thought to be semantically equivalent but cognitively equivalent.[8] But if the sentences φ and $\mathrm{T}t_\varphi$ are cognitively equivalent, it seems that if a rational agent accepts the sentence φ they will also accept the sentence $\mathrm{T}t_\varphi$ and vice versa, that is, from the disquotational perspective we seem justified to assume the following principle:

(TDQ) An agent A sincerely, reflectively, and competently accepts a sentence s (under circumstances properly related to a context c), if and only if, A sincerely, reflectively, and competently accepts the sentence $\mathrm{T}\overline{s}$ (under circumstances properly related to a context c).[9]

Together (DQ), (CDQ), and (TDQ) imply (OS), i.e., the claim that believing and believing-true are semantically equivalent. This suggests that if Field's disquotational perspective on truth is assumed, then the only way to resist (OS) is to reject either (DQ) or (CDQ).

3.1 Rejecting Disquotational Belief Principles?

In way of answering traditional puzzles about belief (CDQ) is frequently rejected. Unfortunately, whilst this may help with answering these puzzles it does not really get us out of the fire in the present case. Although, strictly speaking, we can no longer derive (OS) without assuming (CDQ), (OS) will still be a consequence of (DQ) and (TDQ) for instances φ whenever we accept a sentence *s* expressing φ or accept that s is true. Now, Max clearly accepts 'Goldbach's conjecture is true' and hence by appeal to (DQ) and (TDQ) we obtain that Max believes that Goldbach's conjecture is true if and only if Max believes Goldbach's conjecture, which, we have argued, is intuitively wrong.[10] The moral to draw, it seems, is that if the disquotational perspective is accepted in an unqualified way, then one ought to reject both (DQ) *and* (CDQ), if one wants to resist (OS). However, rejecting (DQ) would be at odds with standard semantics of attitude reports, as the principle is widely accepted in the literature on belief reports. Accordingly, we will refrain from explicitly ruling out (DQ) as a plausible principle. For one, in developing our semantics we wish to remain as neutral as possible with respect to the various theories of attitude reports discussed in the literature: it is not the job of a general semantics to be the arbiter between different philosophical or semantic theories. Rather, such a semantics should make the semantic consequences of the different theories precise. For another, there is a more general reason why one should be wary of rejecting (DQ) in reaction to the derivation of (OS): the principal example we used to argue against (OS) seem to also yield a straightforward argument against (TDQ). Recall the case of Max; for all we know Max would accept 'Goldbach's conjecture is true' but reject 'Every even number >2 is the sum of two prime numbers', that

[8] Cf. Field (1994), pp. 251-52. Field makes a number of qualifications to which we will come back to in due course. See also Künne (2003) and Heck (2020) for a discussion of cognitive equivalence.

[9] $\overline{s}$ is a name of the sentence s.

[10] Admittedly, Max does not accept 'Every even number >2 is the sum of two prime numbers.' and, as we shall discuss below, this yields an argument against (TDQ).

is, Max's acceptance patterns would not be conform with (TDQ). This suggests that the right course of action is to rethink (TDQ) rather than to reject the disquotational belief principles.

3.2 (TDQ) and the Disquotational Case for (OS)

The disquotationalist seems to be left with two options. They can either reject (TDQ) or accept (OS) as a valid principle governing the interaction of truth and belief. Indeed we think that the disquotationalist who holds (TDQ) dear should accept (OS). Of course, this is at odds with Clara's and Max's beliefs but disquotationalists frequently suggest that it is not their aim to capture all reasonable uses of the truth predicate in natural language but only the theoretically useful ones, that is, the disquotational uses of truth.[11] They wish to characterize a truth predicate that can fulfill its theoretical role, i.e. its disquotational function, and to this end it seems required that φ and $\mathrm{T}t_\varphi$ are intersubstitutable salva veritate in contexts like (2) for otherwise, it seems, that (2) would not correctly report Anne's belief.[12] On this view, Max's use of the truth predicate would simply not qualify as a use of the disquotational notion of truth since, according to Field, "*a person can meaningfully apply "true" in the pure disquotational sense only to utterances that he has some understanding of*" (Field, 1994, p. 250). A disquotationalist will hence simply dismiss cases like Max's as irrelevant for his project.[13] They are not legitimate counterexamples to (OS) save (TDQ). On this view, disquotationalist should not flinch and accept (OS)…alas few do.[14]

However, the disquotationalist's dismissal of the counterexamples against (TDQ) and (OS) points to a different possible course of action. In contrast to the diquotationalist position sketched above, one may be more liberal and allow for non-disquotational uses of the truth predicate, that is, uses of the truth predicate for which (TDQ) and (OS) may fail. Arguably, this failure need not concern the disquotationalist, since it is limited to non-disquotational uses of 'true'. The idea would conceive of (TDQ) and (OS) as principles pertaining only to "ideal" or "disquotational" circumstances, that is, circumstances in which—according to the disquotationalist—an agent "has some understanding of" the utterance they deem true. More precisely, if an agent is aware, or understands, which belief is represented, directly or indirectly, by a term t_φ, then a rational agent holds that particular belief, if and only if, they hold t_φ true, that is, in this case they will believe $\mathrm{T}t_\varphi$ if and only if they believe φ. Another way of putting this idea is that (OS) should hold for an instance t_φ, if and only if, the agent thinks about φ and $\mathrm{T}t_\varphi$ in the same way, that is, in this particular case the agent treats the truth predicate transparently. A semantics based on this idea will attribute independent truth-conditions to $\mathrm{B}\varphi$ and $\mathrm{BT}t_\varphi$ that only coincide in case of ideal, disquotational circumstances. Such a semantics should be acceptable to both disquotationalists, and truth theorists that are neither disquota-

[11] See, e.g., Picollo and Schindler (2020) for an endorsement of this view.

[12] A similar point is made by Heck (2020).

[13] The case of Clara is somewhat different. Arguably, the use of the truth predicate is a disquotationally legitimate use. But the disquotationalist would presumable say that 'Superman is strong' does not express that Clark Kent is strong. Rather it expresses whatever Clara qua competent speaker understands 'Superman is strong' to say, or something along these lines.

[14] Indeed, we are not aware of a single disquotationalist who defends (OS). Field (2006) seems to explicitly reject the conclusion. Heck (2020) agrees with our assessment that disquotationalist like Field are committed to (OS).

tionalists nor deflationists:[15] after all $\mathrm{B}\varphi$ and $\mathrm{BT}t_\varphi$ are treated as semantically equivalent if the truth predicate is used disquotationally but the semantics also allows for non-disquotational uses of the truth predicate in which the semantic equivalence breaks down.

3.3 Ways of Believing

In the literature on attitude reports appealing to the way an agent thinks of a given belief, i.e. the way they believe, is a common strategy for explaining allegedly counterintuitive consequences of, roughly, Russellian theories of attitudes. For example, it has been argued that it is possible for a rational agent to believe that φ, while at the same time to believe that $\neg\varphi$ as long as they do not believe φ and $\neg\varphi$ in the same way. There is some disagreement whether the way of believing should be semantically or pragmatically encoded: Naive Russellians such as Soames (1987) typically argue that it should be merely pragmatically encoded whilst contextualists like Crimmins and Perry (1989) embrace the idea that the way of believing ought to be semantically encoded. For example, according to Crimmins and Perry (1989) depending on the way a proposition is believed an agent stands in a different belief-relation to the proposition at stake.[16] Moreover, an agent can believe a proposition φ in one way but believe its negation in another way and, in this case, both $\mathrm{B}\varphi$ and $\mathrm{B}\neg\varphi$ should receive the semantic value true. According to the standard contextualist view the way of believing depends upon an unarticulated constituent of the attitude report and is provided by the context under consideration. In contrast, the semantic picture we are about to propose agrees with the contextualist that the way of believing impacts the semantic evaluation of the attitude report but we do not conceive of it as an unarticulated constituent of the attitude report.[17] Rather, the idea is that the way of believing $\mathrm{BT}t_\varphi$ is determined by the specific representation t_φ. Furthermore, in the absence of further information we are only guaranteed to believe $\mathrm{T}t_\varphi$—assuming we believe it at all—in this specific way, i.e. under the representation t_φ, and this idea will be hardwired into our semantics. In contrast, if no formula of the form $\mathrm{T}t_\psi$ occurs in φ, then φ will be believed in an unspecific way, that is, a way of believing that does not depend on a particular representation of the belief. If we believe a proposition in such an unspecific way there is again no guarantee that we also believe it under a specific representation: $\mathrm{B}\varphi$ and $\mathrm{BT}t_\varphi$ can only be assumed to be equivalent if we are guaranteed that φ and $\mathrm{T}t_\varphi$ are believed in the same way. On our semantics this will be the case, if an agent is aware that by believing that t_φ is true, they commit themselves to believing that φ and vice versa, that is, if an agent is aware that t_φ, directly or indirectly, represents the belief that φ.

[15] We take disquotationalists in contrast to other deflationists to conceive of 'true' as a predicate of sentences or utterances rather than of propositions (cf. Künne, 2003).

[16] Of course, Crimmins and Perry (1989), like other Russellians, conceive of propositions as structured entities, which is at odds with conceiving propositions as sets of possible worlds, states, or situations as customary in PWS. At this point, our comparison only pertains to the idea that the way of believing impacts the attitudinal relation. See Section 6.2 for a discussion of the structured propositions approach.

[17] Or rather we remain neutral whether the way of believing has an impact on the semantic evaluation of the attitude report if it is not explicitly conveyed in reporting the attitude.

4 Semantics for Ways of Believing

In the previous section we argued that the way of believing impacts the semantic value of a belief ascription and that the way of believing depends on the representation t_φ, if $\mathrm{T}t_\varphi$ occurs in the belief context. But this leaves open two alternatives on how t_φ can impact the way of believing: it can either have an impact qua expression of the language or in virtue of what it denotes. Which of the two alternatives one ought to pick, will depend on the objects one takes the truth predicate to apply to. Again there are two options: if, as in the case of disquotational truth, a sentential truth predicate is assumed, the objects of truth are sentences (or utterances) and, as a consequence, the objects of truth will arguably be of a different category than the objects of belief, which typically are thought to be propositions.[18] However, if an a propositional truth predicate is assumed, the objects of truth will be propositions, and the objects of truth and belief will coincide. Of course, depending on the view one opts for, a name t_φ will denote different types of objects, that is, either sentences or propositions. In this paper we makes the simplifying assumption that whether the denotatum of a name t_φ is a sentence or a proposition is not reflected on the linguistic level, that is, we cannot distinguish between names of propositions and names of sentences on purely syntactic or linguistic grounds. Rather we consider it to be a conceptual decision which type of denotata one opts for. More generally, we conceive of a "name" t_φ to be any kind of nominalization that plausibly denotes a sentence or a proposition, that is, t_φ need not be a proper name in the sense of Kripke (1972) but could, e.g., also be a definite description or a that-clause. Similarly, from the perspective of our formal language t_φ will simply be a singular term that denotes the sentence 'φ' or the proposition that φ.

Let us return to the question of whether t_φ impacts the way of believing qua name or in virtue of what it denotes and consider the case of the sentential truth predicate. In this case a term t_φ names a sentence, which, in turn, expresses a proposition. On this view, it seems

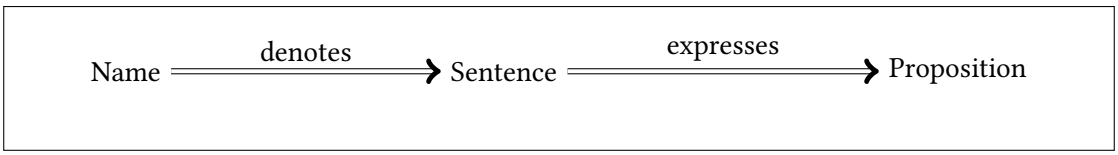

Figure 1: Two-level belief representation

plausible to assume that the reason why an agent may believe, say, that snow is white whilst at the same time not believe that 'Snow is white' is true, is that the agent is not aware that the sentence 'Snow is white' expresses the proposition that snow is white. In other words, from perspective of sentential truth the way we believe is determined by the sentence rather than its name.

We shall adopt the sentential, i.e. disquotational, perspective in formulating our semantics but the view that conceives of the objects of truth as sentences (or utterances) is highly contested. Rather it is often thought that the natural language truth predicate applies to propo-

[18] It is not important for our purpose whether the objects of belief are propositions or some other sort of attitudinal object. The relevant issue is whether the objects of truth and belief coincide or not.

sitions. On this view, t_φ denotes a proposition and believing that $\mathrm{T}t_\varphi$ is not mediated via a sentence that expresses a proposition but depends only on the name t_φ and the proposition itself. Accordingly, if the objects of truth are conceived of as propositions and, at the same

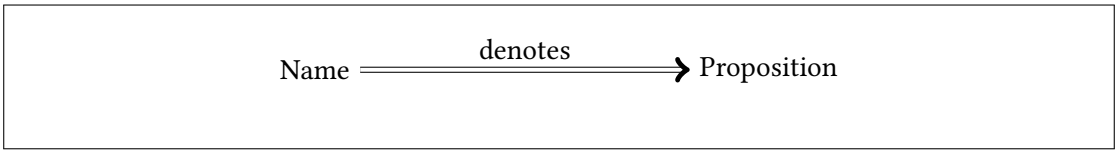

Figure 2: One-level belief representation

time, we wish to maintain the idea that t_φ impacts the way we believe $\mathrm{BT}t_\varphi$, we are forced to accept that t_φ impacts the way of believing qua expression of the language, i.e. qua name, rather than via its semantic contribution. Of course, this idea could also be implemented within the framework of two-level belief representation, but making the way of believing dependent on sentences rather than their names has the neat consequence that the semantics remains fully referential. In contrast, if the name t_φ is allowed to impact the semantic evaluation qua expression, the semantic will no longer be fully referential. So there seems to be at least a prima facie advantage of adopting a framework of two-level belief representation in which the way of believing depends on the denotatum of t_φ rather than t_φ itself.

Independently of whether we opt for a sentential or a propositional truth predicate the question arises how the contribution of the way of believing should be spelled out within the framework of PWS. To this end, it is helpful to coerce the believe relation into the framework of PWS: in PWS an agent a believes the proposition that φ, denoted by $\|\varphi\|$, at a world w iff

$$\forall v(wR_a v \Rightarrow v \in \|\varphi\|),$$

where $\|\varphi\|$—the proposition that φ—is the set of possible worlds in which φ is true and R_a a doxastic accessibility relation. From this definition of the belief-relation one can easily see that the only parameter in the definition is the doxastic accessibility relation. Consequently, if the way of believing is supposed to depend upon t_φ, the term needs to impact the accessibility relation. Indeed, the crucial point of departure of our semantics from standard PWS is that instead of assuming a primitive accessibility relation for every agent we now consider a function that outputs accessibility relations and takes either finite sets of sentences (sentential truth predicate) or finite sets of names (propositional truth predicate) of the language as inputs. We appeal to sets of sentences (terms) rather than to the sentences or names themselves, since we might have several formulas of the form $\mathrm{T}t$ in the belief-context. For example, the truth of $\mathrm{B}(\mathrm{T}t_\varphi \wedge \mathrm{T}s_\psi)$, that is, the way we believe $\mathrm{T}t_\varphi \wedge \mathrm{T}s_\psi$, should depend on both t_φ and s_ψ, that is, on the set $\{t_\varphi, s_\psi\}$.

Before we spell out the formal semantics in some more detail, we need to reconsider the idea that (TDQ) and (OS) should hold in "ideal" circumstances. Following up on our remarks in Section 3 we take "ideal" or "disquotational" to be circumstances in which the agent is *aware* which proposition the sentence denoted by t_φ expresses (the name t_φ denotes). In this case their belief that $\mathrm{T}t_\varphi$ will be insensitive to the particular way of believing related to t_φ and coincide

with the way the agent believes that φ, that is, relative to t_φ the truth predicate behaves transparently in the relevant belief-context. On our semantics, the information which sentences (names) an agent is aware of needs to be provided externally, i.e., it is not possible to compute the Awareness set on the basis of the information provided within the semantics. In this respect our semantics resembles classical Awareness semantics (cf. Fagin et al., 1995, Ch. 9.5), where an externally provided Awareness set controls the transition from idealized belief to non-idealized belief. It is perhaps best to view the Awareness set to be retrieved from the information provided by the common ground relevant to the particular belief report but we shall leave this issue open for the purpose of our formal semantics.

4.1 Formal Semantics

In this section we make our heuristic remarks precise and introduce a formal semantics for ways of believing for a language containing the belief operator B and the truth predicate T.[19] As we remarked in the previous section we shall assume a sentential truth predicate, that is, we shall develop a semantics for two-level belief representation. We wish to keep our semantics as general as possible. For this reason we appeal to an inner/outer domain semantics (cf., e.g., Garson, 2001), that is, we allow for a universe of discourse U, which comprises the domain of quantification, which is allowed to vary from world to world. We also allow for terms of the language to denote non-rigidly, as long as these terms are not expressions of the language of the syntax theory. The interpretation of the syntax theory, that is the syntactic vocabulary, will remain constant across worlds. In contrast to more customary formulations the syntax theory will not carry any explicit ontological commitments, as we shall not require the expressions of the language to be in the domain of quantification at each world. Rather any potential commitment should be considered implicit and as a *sine qua non* condition of the theoretical framework; it is a different kind of commitment than the one we engage in when talking about, say, elephants. The question how this type commitment is to be understood is left open but will obviously depend on one's interpretation of the universe U. To avoid explicit commitment to an ontology of expressions the language of syntax is conceived of as a quantifier-free language along the lines of certain formulations of PRA. However, our syntax theory and language need not be an arithmetical language where the syntactic operations operate on codes of expressions, i.e., natural numbers, but could—perhaps preferably—be a syntax theory that is operates directly over an ontology of expressions (see, e.g., Halbach and Leigh, 2020).

Definition 1 (Universe, Language)**.** *Let* $\mathrm{O} \neq \emptyset$ *be the set of non-syntactic objects relevant to the discourse under consideration. Let* $\mathcal{L}_\mathrm{O}$ *be the language of a syntax theory such that* $\mathcal{L}_\mathrm{O}$

- *contains names* $o_1, o_2, \ldots$ *for all member of* O*;*
- *contains the logical constants* $\neg$ *and* $\wedge$ *and the identity predicate but is quantifier free;*
- *contains names for all expressions of* $\mathcal{L}_\mathrm{B}$ *(cf. below) and function symbols for all primitive recursive syntactic operations of* $\mathcal{L}_\mathrm{B}$*.*

[19] To keep the presentation as concise as possible we only allow for one agent—this allows us to omit an index for B and simplifies some of our definitions, yet nothing hinges on this simplifying assumption.

Let $\mathrm{U} := \mathrm{O} \cup \mathsf{Expr}_{\mathcal{L}_\mathrm{B}}$ *be the universe of discourse for* $\mathcal{L}_\mathrm{B}$ *where* $\mathsf{Expr}_{\mathcal{L}_\mathrm{B}}$ *is the set of all expressions of* $\mathcal{L}_\mathrm{B}$. $\mathcal{L}_\mathrm{B}$ *extends* $\mathcal{L}_\mathrm{O}$ *by a countable number of individual constants* $c_1, c_2, \ldots$*; n-ary predicate symbols* $P^n_1, P^n_2, \ldots$*; the belief operator* B*; the truth predicate* T*; (possibly) the Awareness predicate* A *and the universal quantifier* $\forall$. *Other logical symbols are used merely as abbreviations. The syntax of* $\mathcal{L}_\mathrm{B}$ *is given by*

$$\varphi ::= \psi \mid t_i = t_j \mid P^n t_1, \ldots, t_n \mid \mathrm{A}t \mid \mathrm{T}t \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathrm{B}\varphi \mid \forall x \varphi$$

with $\psi \in \mathsf{Frml}_{\mathcal{L}_\mathrm{O}}$ *and* $t_1, \ldots, t_n, t_i, t_j \in \mathsf{Term}_{\mathcal{L}_\mathrm{B}}$ *where* $\mathsf{Frml}_{\mathcal{L}_\mathrm{O}}$ *(*$\mathsf{Term}_{\mathcal{L}_\mathrm{B}}$*) is the set of formulas (terms) of* $\mathcal{L}_\mathrm{B}$.

The definition implies that the cardinality of the language will depend on the set of contingent objects O the language is intended to talk about. In particular, if O is uncountable, then $\mathcal{L}_\mathrm{O}$ and $\mathcal{L}_\mathrm{B}$ will also be uncountable.

With the details of the languages $\mathcal{L}_\mathrm{O}$ and $\mathcal{L}_\mathrm{B}$ in place, we need to specify their semantic interpretation. To this end we introduce the notion of a belief frame. Models for $\mathcal{L}_\mathrm{B}$ will be defined relative to such a belief frame.

Definition 2 (Belief frame)**.** *A belief frame* F *is a tuple* $\langle \mathrm{U}, W, H, D \rangle$ *where* U *is the universe of discourse,* $W \neq \emptyset$ *is a set of worlds, and* $D : W \longrightarrow \mathcal{P}(\mathrm{U})$ *for all* $w \in W$ *is a function that assigns the domain of quantification relative to a world* w *such that* $\mathrm{O} \subseteq \bigcup_{w \in W} D(w)$*. Finally,* $H : \mathcal{P}(\mathsf{Sent}_{\mathcal{L}_\mathrm{B}}) \longrightarrow W \times W$ *is a function that generates a serial (right-unbounded) doxastic accessibility relation relative to a set of sentences.*[20]

It's worth noting that every non-syntactic object needs to exist at some world. No condition of this kind is imposed on the syntactic objects, that is, the expressions of $\mathcal{L}_\mathrm{O}$. These may live in U without "coming into existence" at any world.

We now define an interpretation over a belief frame F. As mentioned at the beginning of this section the interpretation will act rigidly on $\mathcal{L}_\mathrm{O}$ but is allowed to vary from world to world for the remaining vocabulary of $\mathcal{L}_\mathrm{B}$. As we shall see later, not every interpretation over F gives rise to a belief model since only specific interpretations will ultimately be deemed adequate.

Definition 3 (Interpretation)**.** *Let* F *be a belief-frame. An interpretation* I *is a function that assigns an interpretation to the nonlogical vocabulary of* $\mathcal{L}^-_\mathrm{B}$ *(*$\mathcal{L}_\mathrm{B}$ *without the truth predicate) relative to a possible world such that for all* $w, v \in W$

(i) for all individual constants $k \in \mathcal{L}_\mathrm{O}$, $I(k, w) = I(k, v) \in \mathrm{U}$*; for all individual constants* $k \in (\mathcal{L}_\mathrm{B} - \mathcal{L}_\mathrm{O}), I(k, w) \in \mathrm{O}$*;*

(ii) for all function symbols f^n *of* $\mathcal{L}_\mathrm{O}$, $I(f, w) : \mathrm{U}^n \longrightarrow \mathrm{U}$ *and for all* $u_1, \ldots, u_n \in \mathrm{U}$

$$I(f^n, w)(u_1, \ldots, u_n) = I(f^n, v)(u_1, \ldots, u_n);$$

(iii) $\mathrm{U} - \mathsf{Sent}_{\mathcal{L}_\mathrm{B}} \subseteq I(\mathrm{A}, w) \subseteq \mathrm{U}$*;*

[20] Of course, further properties of the doxastic accessibility relation could be imposed. Here, we take seriality to be a minimal condition of a doxastic accessibility relation. However, nothing we say in this paper will depend on the properties of the doxastic accessibility relation.

(iv) If P^n is a predicate constant of $\mathcal{L}_{\mathrm{B}}$, then $I(P^n, w) \subseteq \mathrm{U}^n \times \mathrm{U}^n$; for $P^n \in (\mathcal{L}_{\mathrm{B}} - \mathcal{L}_{\mathrm{O}})$

- *if $u_i \in \mathsf{Expr}_{\mathcal{L}_{\mathrm{B}}}$, for some $1 \leq i \leq n$, then for all $e \in \mathsf{Expr}_{\mathcal{L}_{\mathrm{B}}}$*

$$\langle u_1, \ldots, u_i, \ldots, u_n \rangle \in I(P^n, w) \Longleftrightarrow \langle u_1, \ldots, e, \ldots, u_n \rangle \in I(P^n, w);$$

for $P^n \in \mathcal{L}_{\mathrm{O}}$,

- *$I(P^n, w) = I(P^n, v) = \langle X, Y \rangle$ with $Y = \mathrm{U} - X$;*

The definition guarantees that the vocabulary of $\mathcal{L}_{\mathrm{O}}$ is interpreted rigidly and that the expressions of $\mathcal{L}_{\mathrm{B}} - \mathcal{L}_{\mathrm{O}}$ do not discriminate between syntactic objects but only objects of O. It is worth highlighting that Definition 3 does not guarantee that sentence denoting singular terms will always be rigid designators: we only know that if the term is an $\mathcal{L}_{\mathrm{O}}$-expression, then it will be a rigid designator.[21] We will examine the issue more closely in Section 6.1 but until then, to keep things as simple as possible, we conceive of all sentence denoting singular terms as rigid designators. Finally, condition (iii) of Definition 3 specifies that all objects in U that are not sentences will always be in the extension of the Awareness predicate, that is, the interpretation of the Awareness predicate at a world can only vary with respect to the sentences in its extension. This may seem awkward for one might wonder what it means to be aware of some non-syntactic object. However, the point is that we intend the Awareness predicate to discriminate only between different sentences rather than arbitrary objects.[22] As explained at the beginning of Section 4, in the semantics of two-level belief representation we adopted the doxastic accessibility relation under consideration will depend on which sentences an agent is aware of. As laid out in Remark 13 below, if we were to work in a system of one-level belief representation instead, the interpretation of the Awareness predicate should discriminate between names of sentences, or more correctly, names of propositions.

Definition 3 does not guarantee that the syntactic operations and expressions are interpreted in a desirable way, that is, that they are interpreted as the syntactic operations and the expressions they intend to denote. Interpretations that guarantee such an intended interpretation will be called adequate and give rise to a belief model.

Definition 4 (Adequate Interpretation, Belief Model, Assignment)**.** *Let* (U, J) *be the standard model (of the syntax theory of) of $\mathcal{L}_{\mathrm{O}}$. We call an interpretation function I adequate iff for all non-logical constants (individual, function, predicate constants) $\zeta \in \mathcal{L}_{\mathrm{O}}$ and all $w \in W$, $\mathsf{J}(\zeta) = I(\zeta, w)$. If I is an adequate interpretation, then $M = (F, I)$ is called a belief model. An assignment $\beta : Var_{\mathcal{L}_{\mathrm{O}}} \longrightarrow \mathrm{U}$ assigns to each variable an object in* U.

We now turn to the interpretation of the truth predicate, which is provided by an evaluation function.

[21] Suppose the function symbol q represents a function on U that, similar to the num-function, if applied to some element of U outputs the "standard" name of the object. Then $\mathsf{q}(c_1)^\frown \doteq^\frown \mathsf{q}(c_2)$ is a name of a sentence. But since the interpretation of c_1 and c_2 may change from world to world, $\mathsf{q}(c_1)^\frown \doteq^\frown \mathsf{q}(c_2)$ may denote, say, the sentence $o_1 = o_2$ at world w but the sentence $o_3 = o_5$ at world v.

[22] Indeed we could have also defined the interpretation of the Awareness predicate to be a set of sentences only and modify Definition 6 below accordingly.

Definition 5 (Evaluation function)**.** *Let F be a belief frame. An evaluation function relative to F is a function $f : W \times W \longrightarrow \mathcal{P}(\mathsf{Sent}_{\mathcal{L}_{\mathrm{B}}})$ that assigns to each pair of worlds a set of sentences—the interpretation of the truth predicate. The set of all evaluation functions relative to a frame F is denoted by Val_F.*

Next we introduce the index set of a formula φ, which serves as the input to the H function that outputs an accessibility relation. Intuitively the index set of φ consists of all sentences ψ such that $\mathrm{T}t_\psi$ is a subformula of φ and where the agent is not aware which proposition the sentence ψ expresses, i.e., the set of all those representations occurring (explicitly and implicitly) in φ which are not transparent to the agent. It is in the definition of the index set where the differences between systems of one-level and two-level belief representation become most apparent. While in the present case of two-level belief representation the index set will consist of a set of sentences, it will be a set of names (of propositions) in the case of one-level belief representation, namely, the set of those names t_ψ such that the agent is not aware of the proposition t_ψ denotes.

Definition 6 (Index set)**.** *Let φ be a formula of $\mathcal{L}_{\mathrm{B}}$, $t^{M,w}[\beta]$ be the interpretation of a term t in M at w relative to a variable assignment β. Then the operation $\varphi^\beta_w : \mathsf{ON} \longrightarrow \mathcal{P}(\mathsf{Sent}_{\mathcal{L}_{\mathrm{B}}})$ is defined by the following recursion for $\alpha, \kappa \in \mathsf{ON}$:*

$$
\begin{aligned}
\varphi^\beta_w(0) &:= \emptyset \\
\varphi^\beta_w(\alpha+1) &:= \begin{cases}
\emptyset, & \text{if } \varphi \in \mathcal{L}^-_{\mathrm{B}} \text{ or } (\varphi \doteq \mathrm{T}t \,\&\, t^{M,w}[\beta] \notin \mathsf{Sent}_{\mathcal{L}_{\mathrm{B}}}); \\
\{\psi\}, & \text{if } \varphi \doteq \mathrm{T}t \,\&\, t^{M,w}[\beta] = \psi \in \mathcal{L}^-_{\mathrm{B}} \,\&\, \psi \notin I(w, \mathrm{A}); \\
\psi^\beta_w(\alpha), & \text{if } \varphi \doteq \mathrm{T}t \,\&\, t^{M,w}[\beta] = \psi \in I(w, \mathrm{A}); \\
\{\psi\} \cup \psi^\beta_w(\alpha), & \text{if } \varphi \doteq \mathrm{T}t \,\&\, t^{M,w}[\beta] = \psi \notin I(w, \mathrm{A}); \\
\psi^\beta_w(\alpha), & \text{if } \varphi \doteq \neg\psi \text{ or } \varphi \doteq \mathrm{B}\psi; \\
\psi^\beta_w(\alpha) \cup \chi^\beta_w(\alpha), & \text{if } \varphi \doteq \psi \wedge \chi; \\
\bigcup_{d \in D(w)} \psi(v)^{\beta(v:d)}_w(\alpha), & \text{if } \varphi \doteq \forall v \psi.
\end{cases} \\
\varphi^\beta_w(\mu) &:= \bigcup_{\alpha < \mu} \varphi^\beta_w(\alpha), \text{ if } \mu \text{ is limit.}
\end{aligned}
$$

Let $\kappa \in \mathsf{ON}$ be such that $\varphi^\beta_w(\kappa) = \varphi^\beta_w(\alpha)$ for all $\alpha \geq \kappa$.[23] *Then $\varphi^\beta_w := \varphi^\beta_w(\kappa)$ is called the* INDEX SET *of φ relative to a belief model $\mathcal{M}$, a world w and an assignment function β.*

We now define truth in a model at a world in a belief-model relative to an evaluation function f. Indeed, as already implicit in Definition 5 we define truth in a model relative to a pair of worlds (w, v) where w is the world in which we evaluate the given formula and v is the world relative to which the index set of the formula is defined. The reason for the two-dimensional interpretation is that we want to evaluate a formula relative to the interpretation of the Awareness predicate at the initial world of evaluation rather than one of its doxastic alternatives which may be relevant for evaluating one of its subformulae at a later stage of the semantic computation. For example, consider the formula $\mathrm{BBT}t_\varphi$. In evaluating the formula at

[23] Since φ^β_w is monotone, we know that such a κ exists.

w, that is (w, w), we first check whether $\mathrm{BT}t_\varphi$ is true in all doxastic alternatives v of w given the accessibility relation generated by the index set of $\mathrm{BT}t_\varphi$ at w. In the next step, we wish to consider whether $\mathrm{T}t_\varphi$ is true at all doxastic alternatives of v relative to the index set of $\mathrm{T}t_\varphi$ at w—rather than the index set of $\mathrm{T}t_\varphi$ at v—but without appealing to a two-dimensional interpretation there is no way we can retrieve the starting point of our semantic computation. Again, the point is that what matters in the semantic evaluation of the $\mathrm{BBT}t_\varphi$ is what the agent is aware of at world w rather than at a doxastic alternative v. Moreover, the index set of $\mathrm{T}t_\varphi$ relative to w may be different to the index set of $\mathrm{T}t_\varphi$ relative v to since the interpretation of the Awareness predicate may change and, as a consequence, $\mathrm{BT}t_\varphi$ may be true at v, say, relative to the index set of $\mathrm{T}t_\varphi$ at w but false relative to its index set at v.

Definition 7 (Strong Kleene Truth in a Belief Model)**.** *Let F be a belief frame, $f \in \mathsf{Val}_F$ and β a variable assignment. We define the notion of truth in a belief model M relative to the evaluation function f and the assignment β at a world-pair (w, v) according to the strong Kleene scheme for formulas φ of $\mathcal{L}_\mathsf{B}$ ($M, (w, v) \vDash^f_{\mathsf{sk}} \varphi[\beta]$) by an induction on the positive complexity of φ:*

(i) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} s = t[\beta] \iff s^{M,w}[\beta] = t^{M,w}[\beta]$

(ii) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg(s = t)[\beta] \iff s^{M,w}[\beta] \neq t^{M,w}[\beta]$

(iii) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} P^n t_1, \dots, t_n s = t[\beta] \iff \left\langle t_1^{\mathcal{M},ww}[\beta], \dots, t_n^{\mathcal{M},w}[\beta] \right\rangle \in I(P^n, w)^+;$

(iv) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg P^n t_1, \dots, t_n[\beta] \iff \left\langle t_1^{\mathcal{M},ww}[\beta], \dots, t_n^{\mathcal{M},w}[\beta] \right\rangle \in I(P^n, w)^-;$

(v) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \mathrm{A}t[\beta] \iff t^{M,w}[\beta] \in I(\mathrm{A}, w)$

(vi) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg\mathrm{A}t[\beta] \iff t^{M,w}[\beta] \notin I(\mathrm{A}, w)$

(vii) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \mathrm{T}t[\beta] \iff t^{M,w}[\beta] \in f(w, v)$

$(viii)$ $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg\mathrm{T}t[\beta] \iff (\dot{\neg}t)^{M,w}[\beta] \in f(w, v)$ *or* $t^{M,w}[\beta] \notin \mathsf{Sent}_{\mathcal{L}_\mathrm{B}}$,

(ix) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg\neg\psi[\beta] \iff M, (w, v) \vDash^f_{\mathsf{sk}} \psi[\beta]$

(x) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \psi \wedge \chi[\beta] \iff (M, (w, v) \vDash^f_{\mathsf{sk}} \psi[\beta]$ *and* $M, (w, v) \vDash^f_{\mathsf{sk}} \chi[\beta])$

(xi) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg(\psi \wedge \chi)[\beta] \iff (M, (w, v) \vDash^f_{\mathsf{sk}} \neg\psi[\beta]$ *or* $M, (w, v) \vDash^f_{\mathsf{sk}} \neg\chi[\beta])$

(xii) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \forall x\psi[\beta] \iff$ *for all* $d \in D(w)(M, (w, v) \vDash^f_{\mathsf{sk}} \psi[\beta(x : d)])$

$(xiii)$ $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg\forall x\psi[\beta] \iff$ *there is an* $d \in D(w)(M, (w, v) \vDash^f_{\mathsf{sk}} \neg\psi[\beta(x : d)])$

(xiv) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \mathrm{B}\psi[\beta] \iff \forall z(H_{\psi^\beta_v} wz \Rightarrow M, (z, v) \vDash^f_{\mathsf{sk}} \psi[\beta])$

(xv) $\quad M, (w, v) \vDash^f_{\mathsf{sk}} \neg\mathrm{B}\psi[\beta] \iff \exists z(H_{\psi^\beta_v} wz \,\&\, M, (z, v) \vDash^f_{\mathsf{sk}} \neg\psi[\beta])$

If a formula φ is true in the belief model M at (w, w) relative to the evaluation function f and assignment β, we write $M, w \vDash^f_{\mathsf{sk}} \varphi[\beta]$ and say that φ is true in the belief model M at w relative to the evaluation function f; if φ is true in M relative to the evaluation function f and the assignment β for all $w \in W(M, w \vDash^f_{\mathsf{sk}} \varphi[\beta])$, we write $M \vDash^f_{\mathsf{sk}} \varphi$ and say that φ is true in M relative to the evaluation function f and the assignment β. We say that φ is true in F under a given evaluation function f and assignment β ($F \vDash^f_{\mathsf{sk}} \varphi[\beta]$)) iff for all belief-models M on F, $M \vDash^f_{\mathsf{sk}} \varphi[\beta]$. In general, we drop reference to an assignment β if the formula is true relative to all assignments.

This concludes the specifications of the semantics for truth and belief. However, it remains to be shown that adequate interpretations of the truth predicate can be constructed over arbitrary belief models, that is, that there are evaluation functions f such that (TSW) holds at each world. Fortunately, this can be shown by running a standard Kripkean construction. To this end, we first define an ordering on Val_F.

Definition 8 (Ordering). *Let $f, g \in \mathsf{Val}_F$. We set $f \leq g$ iff $f(w, v) \subseteq g(w, v)$ for all $w, v \in W$.*

It is not difficult to see that $\models_{\mathsf{sk}}$ is a monotone evaluation scheme relative to the ordering $\leq$.

Fact 9 (Monotonicity). *For $f, g \in \mathsf{Val}_F$ and for all $\varphi \in \mathcal{L}_{\mathsf{B}}$*

$$f \leq g \Longrightarrow \forall w, v \in W(M, (w, v) \models^{f}_{\mathsf{sk}} \varphi \Longrightarrow M, (w, v) \models^{g}_{\mathsf{sk}} \varphi)$$

Fact 9 guarantees the existence of fixed points for arbitrary belief frames F, that is, evaluation functions f such that for all $\varphi \in \mathsf{Sent}_{\mathcal{L}_{\mathsf{B}}}$ and all $w, v \in W$

$$F, (w, v) \models^{f}_{\mathsf{sk}} \varphi \Longleftrightarrow \varphi \in f(w, v).$$

To construct such a fixed point we define a Kripke jump in the customary fashion:

Definition 10 (Kripke Jump). *Let F be a frame and Val_F the set of evaluation functions relative to F and M a belief model Then $\mathsf{SK}_{\mathrm{B}} : \mathsf{Val}_F \longrightarrow \mathsf{Val}_F$ is an operation such that*

$$[\mathsf{SK}_{\mathrm{B}}(f)](w, v) := \{\varphi \mid M, (w, v) \models^{f}_{\mathsf{sk}} \varphi\}.$$

The existence of fixed points of SK_{B} then follows by the usual arguments (see, e.g., Kripke, 1975; Visser, 1984).

Proposition 11 (Fixed points). *Let F be a frame and M a belief model on F. Then there exists an evaluation function $f \in \mathsf{Val}_F$ such that*

$$\mathsf{SK}_{\mathrm{B}}(f) = f.$$

As an immediate corollary of Proposition 11, f provides an adequate interpretation of the truth predicate relative to every world $w \in W$:

Corollary 12 (Truth model). *Let F be a frame and $f \in \mathsf{Val}_F$ a fixed point of SK_{B}. Then for all $w, v \in W$ and $\varphi \in \mathcal{L}_{\mathrm{B}}$*

$$M, (w, v) \models^{f}_{\mathsf{sk}} \mathrm{T}t_{\varphi} \Longleftrightarrow M, (w, v) \models^{f}_{\mathsf{sk}} \varphi.$$

This concludes out presentation of our semantics of truth and belief within the framework of two-level belief representation. Before we turn to some of the consequences of the semantics we sketch out a semantics for one-level belief representation,

Remark 13 (One-level Belief Representation)**.** *The semantics was developed with a two-level belief representation in mind, that is, the language contains names of sentences, which in turn express propositions (or some other attidunal object). A semantics of one-level belief representation can be obtained by implementing the following changes:*

- *the syntax theory needs to be conceived of as, or replaced by, a theory of structured propositions;*[24]
- *the arguments of H need to be sets of names of propositions rather than propositions;*
- *the interpretation of* A *will be a set of names rather than a set of their denotata;*
- *the set φ^{β}_{w} will also be a set of names; moreover when $\varphi \doteq \mathrm{T}t$, where t denotes a proposition ψ relative to β and $t(\overline{\beta(x)}/x) \notin I(w, \mathrm{A})$ one should probably set $\varphi^{\beta}_{w} := \{t(\overline{\beta(x)}/x)\}$ rather than $\psi^{\beta}_{w} \cup \{t(\overline{\beta(x)}/x)\}$ as suggested by Definition 6. The reason is that in this case if $t(\overline{\beta(x)}/x) \notin I(w, \mathrm{A})$ the agent will have no grasp of the proposition denoted by t and, in particular, they will not be aware of the belief-representation, i.e. names of propositions, the proposition appeals to. In contrast, in the case of two-level belief representations the agent is under no illusion what sentence a particular name refers to: they are only not aware which particular proposition the sentence expresses.*

4.2 Formal Consequences

The semantics behaves as intended in the sense that (OS) is not generally true at a world w in a belief model, indeed both directions of (OS) fail:

$$(\nRightarrow) \qquad M, w \models^{f}_{\mathsf{sk}} \mathrm{B}\varphi \nRightarrow M, w \models^{f}_{\mathsf{sk}} \mathrm{BT}t_{\varphi}$$

$$(\nLeftarrow) \qquad M, w \models^{f}_{\mathsf{sk}} \mathrm{B}\varphi \nLeftarrow M, w \models^{f}_{\mathsf{sk}} \mathrm{BT}t_{\varphi}.$$

However, as intended, (OS) holds under idealization conditions, that is, if $t^{M,w}_{\varphi}[\beta] \in I(\mathrm{A}, w)$ then

$$M, w \models^{f}_{\mathsf{sk}} \mathrm{B}\varphi[\beta] \iff M, w \models^{f}_{\mathsf{sk}} \mathrm{BT}t_{\varphi}[\beta].$$

This follows from the fact that if $t^{M,w}_{\varphi}[\beta] \in I(w, \mathrm{A})$, $\varphi^{\beta}_{w} = (\mathrm{T}t_{\varphi})^{\beta}_{w}$.

For the T-free fragment of the language the B-operator behaves like a standard modal operator in possible world semantics, that is, for the T-free fragment B is a modal operator of a normal modal logic. This is in stark contrast to the behavior of B if applied to sentences in which T occurs. In this case B is a non-normal and indeed a hyperintensional and non-algebraic operator, that is, let $\|\varphi\|^{f}_{M} := \{w \in W \mid M, w \models^{f}_{\mathsf{sk}} \varphi\}$, then

$$\|\varphi\|^{f}_{M} = \|\psi\|^{f}_{M} \Longrightarrow \|\mathrm{B}\varphi\|^{f}_{M} = \|\mathrm{B}\psi\|^{f}_{M}$$

[24] Admittedly, this sits ill with PWS where propositions are conceived of as sets of possible worlds. We shall ignore this issue for the purpose of the formal semantics but see Section 6.2.

does not generally hold if the truth predicate occurs in either φ or ψ. In this respect our semantics is similar to semantics of non-idealized belief such as classical Awareness semantics or Impossible world semantics. But while in these semantics whether a formula $\mathrm{B}\varphi$ or $\mathrm{BT}t_\varphi$ is satisfied at some world would depend solely on the Awareness operator or the Impossible world, our semantics provides independent truth conditions for the two formulas, which happen to coincide if t_φ is in the interpretation of the Awareness predicate in the world under consideration.

One consequence of the non-normality of B is that the operator does not generally commute with conjunction at a world w if either of the conjuncts has a subformula of the form $\mathrm{T}t$ such that $t^{M,w}[\beta] \notin I(\mathrm{A}, w)$:

$$M, w \models^f_{\mathsf{sk}} \mathrm{B}(\psi \wedge \chi) \not\Leftrightarrow M, w \models^f_{\mathsf{sk}} \mathrm{B}\psi \wedge \mathrm{B}\chi.^{25}$$

The reason is that while we may believe ψ and ϕ in some way we might not believe them in the same way. It is precisely for this reason that it is possible for both $\mathrm{B}\varphi$ and $\mathrm{B}\neg\mathrm{T}t_\varphi$ to be true at a world w. However, it is impossible for $\mathrm{B}(\varphi \wedge \neg\mathrm{T}t_\varphi)$ to be true at any world w since this would imply that we believe φ and $\neg\mathrm{T}t_\varphi$ in the same world, which contradicts Corollary 12. This kind of phenomena also arises in contextualist theories of belief, where it is possible for Clara to believe that Superman is strong and to believe that Clark Kent is not strong but impossible for her to believe that Superman is strong and that Clark Kent is not strong in the same way.

Moreover, due to the same phenomenon an agent will believe the T-scheme for grounded sentences despite the fact that (OS) will not generally hold at a world even for such sentences, that is, if $F \models^f_{\mathsf{sk}} \mathrm{T}\dot{\neg}t_\varphi \leftrightarrow \neg\mathrm{T}t_\varphi$, then

$$F \models^f_{\mathsf{sk}} \mathrm{B}(\mathrm{T}t_\varphi \leftrightarrow \varphi).^{26}$$

At first glance the semantics seems to get a lot of things right but how does it apply to the case of Max and Clara respectively. Does it yield the correct semantic explanations and predictions?

5 Taking Stock: Anne, Clara and Max

We started the paper by observing that standard PWS for belief combined with the idea that semantic states-of-affairs supervene on non-semantic states-of-affairs leads to the undesirable consequence that believing and believing-true are semantically equivalent. Such a semantics,

[25] Similarly, disjunction introduction in the scope of B fails, i.e.,

$$M, w \models^f_{\mathsf{sk}} \mathrm{B}\psi \not\Rightarrow M, w \models^f_{\mathsf{sk}} \mathrm{B}(\psi \vee \chi).$$

[26] We take this to be a neat feature of our semantics but, to be sure, it is not unproblematic. Because of basically the same phenomenon $\mathrm{B}(\mathrm{T}t_\varphi \vee \varphi)$ and $\mathrm{BT}t_\varphi$ will be equivalent on our semantics, which seems problematic if you think of a case like Max's. Thanks to Ollie Tatton-Brown for raising this issue.

we argued, yields counterintuitive accounts and predictions in the case of Clara and Max respectively. The moral we drew from this observation was that the semantic value of a belief report depends on the way an agent believes—akin to an idea prominent in contextualist theories of belief reports. In particular, we argued that the way we believe φ will depend on the syntactic representations of beliefs occurring in φ and that unless an agent is aware of t_φ, an agent will believe φ and $\mathrm{T}t_\varphi$ in different ways: they may believe a proposition in one way but fail to believe it in another way. On this semantic picture, believing and believing-true are no longer semantically equivalent and we obtain a neat semantic explanation of this fact. But what precisely happens in the cases of Max and Clara respectively? Does the new semantics yield correct semantic predications?

Let us start with Clara and assume (DQ) is correct, that is, that Clara believes that Clark Kent is strong. In this case, it seems, we are compelled to accepting that Clara is not fully aware of the sentence 'Clark Kent is strong' for it is part of the story that Clara does not believe the proposition in a 'Clark Kent is strong'-way, indeed Clara does not seem fully aware of what proposition 'Clark Kent is strong' expresses. (OS) cannot be applied, that is, Clara does not believe that 'Clark Kent is strong' is true. Notice, however, that our semantics does not commit us to accepting (DQ), i.e., to accept that Clara believes that Clark Kent is strong because Clara accepts the sentence 'Superman is strong' and, according to our story, also believes that 'Superman is strong' is true: it is perfectly acceptable according to our semantics for Clara to believe that Clark Kent is strong in a 'Superman is strong'-way but not in a representation independent way, indeed, this seems to be a rather plausible view. But no matter one's position in this respect, the semantics provides the flexibility of reporting Clara's beliefs appropriately.

What about Max? We have already seen that Max's use of the truth predicate does not qualify as a disquotational use and (OS) should not be applicable in this context. Admittedly, even though we have assumed a two-level belief representation, in Max's case it would seem more appropriate to make the way of believing dependent on the name 'Goldbach's conjecture' rather than the sentence it denotes. After all, Max may be in no doubt about which proposition is expressed by a particular sentence expressing Goldbach's conjecture but, according to our story, he does not seem to be aware of the sentence the term 'Goldbach's conjecture' denotes.[27] Now, independently of the particulars of our semantics it seems clear that Max is not aware of the proposition represented, be it directly or indirectly, by the name 'Goldbach's conjecture'. Max does not believe that every even number >2 is the sum of two primes in a representation independent way. Rather the only way Max believes the proposition is in a 'Goldbach's conjecture'-way of believing—(OS) cannot be applied, that is, our semantics gives the correct semantic assessment: 'Max believes that Goldbachs conjecture is true.' and 'Max believes that every even number >2 is the sum of two primes.' are not semantically equivalent.

It seems that our semantics yields the right outcome in Clara's and Max's case where believing and believing-true are not semantically equivalent. But what about disquotational uses of the truth predicate: can our semantics accommodate such uses as we have claimed at the beginning of Section 4? We have seen that the disquotationalist will argue that (2) correctly reports Anne's beliefs but that this requires (OS). In our semantics, application of (OS) is only licensed

[27] This suggests that the uniform treatment of sentence-denoting singular terms in our semantics may be too coarse grained and that we should treat different types of singular terms differently.

if Anne is, for every axiom of Euclidean geometry, aware of at least one sentence expressing the axiom qua proposition. From the disquotational perspective this is arguably an acceptable assumption: the disquotationalist's claim is not that Anne would necessarily report her belief in this way. Rather, the claim is that Anne's belief is correctly reported by (2) if an external, observational perspective is assumed. In reporting Anne's belief by (2) the disquotationalist stipulates the transparency of the truth predicate, that is, they stipulate Anne's awareness of the relevant sentences and it is precisely against the background of this stipulation that the disquotationalist's report of Anne's beliefs is acceptable.

Summing up, our semantics provides an intuitive explanation of why believing and believing-true ought to be semantically differentiated, which neatly applies to the cases of Max and Clara. Moreover, it is sufficiently flexible to accommodate disquotational uses of the truth predicate, that is, uses that treat the truth predicate in a transparent way and for which believing and believing-true turn out to be semantically equivalent. In this respect our semantics should be acceptable to the disquotationalist, although, the disquotationalist will need to grant that there may also be non-disquotational uses of 'true'. However, as for every semantics our semantics also produces some—arguably—counterintuitive consequences and faces certain limitations. To conclude the paper, we discuss some of these limitations point towards some alternative semantic explanations for distinguishing between believing and believing-true.

6 Limitations and Alternative Semantic Explanations

The semantics we presented provides a greater amount of flexibility than standard PWS, which enables the semantics to differentiate between believing and believing-true. But the flexibility of the semantics has its limitations. In this section, we flag two such limitations before outlining some alternative explanations of the semantic difference between believing and believing-true. The first limitation stems from the fact that the way of believing is fully governed by the syntactic information provided by the formula in the scope of the B-operator and the Awareness set; all other contextual information is discarded as irrelevant. The second limitation is inherited from PWS semantics: propositions are still conceived of as sets of possible worlds (or states), which has some at least prima facie undesirable side effects.

To illustrate the first problem recall that in discussing some of the formal consequences of our semantics we noted that B was a non-normal modal operator, if applied to a formula which has a subformula of the form Tt. As a consequence, virtually all logical reasoning will break down in such cases. But frequently agents are perfectly capable of reasoning logically and the semantics should be able to accommodate logical reasoning in such cases. For example, at least at the outset, disjunction introduction in the scope of B seems fine in many cases, i.e.,

$$\frac{\mathrm{B}\varphi}{\mathrm{B}(\varphi \vee \psi)}\,.$$ [28]

But the inference breaks down because the way of believing may change in course of our logical reasoning. While this may happen in some cases, surely, in most circumstances if an agent

[28] It seems particularly unfortunate that disjunction introduction is valid iff ψ does not contain a subformula of the form Tt but invalid otherwise. The reason for this asymmetry is that in the former case ψ does not affect the way of believing while in the latter case it does.

is engaged in reflective, logical thinking, we should expect the way of believing to remain constant throughout the reasoning. This suggests that the way of believing is not fully determined by the syntactic information available, but depends more directly on the context of the belief report. More generally, it seems reasonable to assume that the way of believing will frequently be determined by previous discourse and its common ground.[29] Perhaps then a semantics that explicitly appeals to the way of believing needs to embrace a contextualist approach to attitude reports more fully than we acknowledged in Section 3.

Let's turn to the second limitation. In our semantics a formula φ is true at a world w if and only if $\mathrm{T}t_\varphi$ is true at w. As a consequence, the (possibly) diverging semantic values of $\mathrm{B}\varphi$ and $\mathrm{BT}t_\varphi$ are due to the different range of quantification of the (interpretation of the) B-operator in the two cases rather than the semantic value of the formulas in the scope of B. But this also means that if a formula φ is true (false) in all worlds, then (OS) will be true for φ and every name t_φ of φ: independently of the range of quantification of B, there will simply be no worlds to falsify (verify) φ. $\mathrm{B}\varphi$ and $\mathrm{B}t_\varphi$ will be semantically equivalent. For example, let $\chi := \psi \vee \neg\psi$ where ψ is a sentence of language of syntax $\mathcal{L}_{\mathrm{O}}$, then (OS) holds, i.e., let M be a belief model on an arbitrary frame F, then for all $w \in W$ and names t_χ of χ

$$M, w \models^f_{\mathsf{sk}} \mathrm{B}\chi \iff M, w \models^f_{\mathsf{sk}} \mathrm{BT}t_\chi.$$

This feature clearly highlights the limited flexibility of our semantics and that the semantics inherits some of the conceptual limitations of standard PWS: it is not possible to distinguish between "different" necessarily true (false) propositions.

In devising semantics there is always a trade off between the flexibility of the semantics and the availability of systematic, i.e. non ad hoc, semantic explanations. The question then arises whether we can increase the flexibility of the semantics, that is, to allow for the possibility to block all instances of (OS) whilst at the same time providing or retaining principled semantic explanations. Of course, there is hardly a clear cut answer to the question and where one theorists will push for a more flexible semantics another theorists will invoke pragmatic strategies to account for allegedly counterintuitive consequences. In this paper we shall not enter this kind of debate but point to two alternative strategies for blocking (OS), which may allow for a greater amount of semantic flexibility.

6.1 Non-Rigid Terms and Scope Distinctions

Throughout this paper we have tacitly assumed that names of expressions of $\mathcal{L}_{\mathrm{B}}$ refer rigidly to these expressions, that is, a name t_φ refers to φ in all possible worlds. But this assumption may be questioned, even if one conceives of proper names as rigid designators because we frequently designate sentences or propositions via definite descriptions rather than proper names viz 'the sentence '…'' or 'the proposition that …'. Even if definite description are treated as full-fledged singular terms of the language, rather than incomplete symbols à la Russell (1905), they are generally thought to be flaccid designators and to denote different objects at different worlds. Indeed, as we have mentioned in passing, under certain circumstances our semantics

[29] To some extent this can be accommodated by the fact that the Awareness set is contextually controlled but arguably this will only suffice in cases where the syntactic information remains constant throughout the discourse.

allows for non-rigid sentence denoting singular terms even though we have ignored this aspect of the semantics up to this point. But if non-rigid sentence denoting singular terms are envisaged (OS) would fail because a term t_φ can fail to denote the sentence (proposition) φ in all worlds but the actual world—for all we know it could also denote the sentence (proposition) ψ.[30] Of course, (TSW) would then fail likewise, that is, we would only be guaranteed that $\mathrm{T}t_\varphi$ and φ receive the same semantic value in the actual world but not in any other world. However, this would not imply that semantic states-of-affairs do not supervene on non-semantic states-of-affairs, since (TSW) no longer asserts that 'φ' holds at a world w, if and only if 'the sentence 'φ' (the proposition that φ) is true' holds at w, but possibly, using our previous example, that 'φ' holds at w, if and only if, 'the sentence ψ' (the proposition that φ) is true' holds at w. So, at least prima facie, treating sentence/proposition-denoting singular terms as flaccid designators does not seem to clash with any fundamental semantic principle.

This opens up the possibility of semantically distinguishing between believing and believing-true without appealing to different ways of believing. Rather the distinction arises due to non-rigid designation. However, if sentence/proposition-denoting singular terms are treated as non-rigid designators, the question arises how we are to deal with cases like Anne's, i.e., belief reports like (2) where, at least from a disquotational perspective, appealing to (OS) seems to be legitimate and to yield the correct semantic predictions. The most immediate and plausible strategy to accommodate such cases is to appeal to scope distinctions triggered by the definite description and to distinguish between believing of the denotatum of t_φ that it is true and believing that t_φ is true, that is to distinguish between believing-true *de re* and *de dicto*. While (OS) fails on the de dicto-reading, it seems, or so the argument would go, acceptable on the de re-reading, that is, we would obtain the following version of (OS):

(OSR) $$M, w \vDash^f \mathrm{B}\varphi \iff M, w \vDash^f \langle \lambda x.\mathrm{BT}x \rangle t_\varphi.$$

In Section 5 we argued that the disquotationalist assumes an external, observational position in reporting Anne's belief and that their report need not match the way Anne would report her beliefs. This position goes neatly with the strategy of accommodating belief reports like (2) by focusing on the de re reading of believing-true.

Still one may wonder whether understanding sentence/proposition-denoting singular terms to be flaccid designators is the correct analysis of the cases of Max and Clara. First, if one subscribes to the Millean/Kripkean-doctrine that names are rigid designators, then names of sentences/propositions such as Goldbach's conjecture should be taken to rigidly designate their referent.[31] As a consequence, theorists would need to provide an alternative explanation for why (OS) fails in Max's case, which suggests that alluding to non-rigid designation cannot fully replace the appeal to ways of believing in semantic explanations. Second, and more generally, it is not clear that in the cases of Clara and Max the failure of (OS) is due to a de

[30] Without special restrictions in place a term t_φ could denote, e.g., the symbol '(' or some arbitrary other object in the domain. In our semantics sentence denoting singular terms will always denote sentences but the philosophical question remains of whether this is a plausible assumption once we have allowed for non-rigid designators: what guarantees that non-rigid designators always designate objects of right (syntactic) category?

[31] Arguably, Goldbach's conjecture is a descriptive name but even descriptive names are typically thought to rigidly designate their denotatum. Moreover, the failure of (OS) in Max case does not seem due to the descriptive property conveyed by 'Goldbach's conjecture'.

dicto-reading of believing-true. Arguably, in Max's case the equivalence between believing and believing-true seems to break down on the de re-reading likewise. Similarly, at least focusing on the sentential truth predicate, Clara's case also remains puzzling under a de re-reading of believing-true. Treating sentence-denoting singular terms as flaccid designator then does not seem to resolve the original puzzle: the puzzle seems to be a puzzle about believing-true de re and the cases brought forward against (OS) appear to undermine (OSR) likewise. Of course, allowing for sentences and propositions to be denoted non-rigidly in the semantics—not all sentence/proposition-denoting expressions should be treated uniformly—may be interesting for other reasons, but, as mentioned, these non-rigid terms will not be helpful in semantically distinguishing believing from believing-true de re.

6.2 More Syntax Strategies

The aim of the paper was to provide an adequate semantics for truth and belief. To this end, we appealed to PWS for belief, which, as we have seen, yields counterintuitive consequences unless special precautions are taken. However, many theorists working on the semantics of attitude reports will deem such an approach a non-starter, as it is well-known that PWS provides a very coarse grained—arguably too coarse grained—analysis of attitude reports and semantic content.[32] According to these views, semantic content, i.e. propositions, should not be conceived as sets of worlds or "truth supporting circumstances" (cf. Soames, 1987) but as structured propositions: whilst in PWS the proposition that Mary is smart is conceived of as the set of all those worlds in which Mary is smart, on the structured proposition account it would be the tuple consisting of Mary and the property of being smart, i.e, ⟨Mary, Smartness⟩. In other words, on the structured proposition account a proposition conveys structural or syntactic information that has been extracted or retained from the sentences that express it. On this view, if belief is conceived as a relation between an agent and a proposition or, possibly, the constituents of the proposition, we should not expect (OS) to be true: 'the proposition that it is true that Mary is smart' will express a proposition along the lines of ⟨True, ⟨Mary, Smartness⟩⟩. While ⟨Mary, Smartness⟩ and ⟨True, ⟨Mary, Smartness⟩⟩ will be true at exactly the same worlds, there is no guarantee that an agent will be aware of this fact—they may believe the one without believing the other.

If one agrees with the idea that belief should not be a relation between an agent and a set of possible worlds but rather a relation between an agent and structured propositions, there is no puzzle: an agent may believe φ and not believe $\mathrm{T}t_\varphi$ and vice versa despite the fact that φ and $\mathrm{T}t_\varphi$ have the same truth value in every world. In light of this one may be tempted to abandon PWS. However, upon closer inspection explicit appeal to structured proposition seems inessential for producing correct semantic prediction, as the belief relation of the structured proposition theorist can be recovered in the PWS framework. On this view a structured proposition is just a set of worlds represented in a way that conforms to specific structural constraints. Accordingly, a formula $\mathrm{B}\varphi$ will be true at a world w iff φ is true at all doxastic alternatives of w and the agent believes the proposition that φ qua set of possible worlds under the representation φ. To make this idea precise on would need to say when a formula is true at a world under a specific

[32] See, Soames (1987); King (2013) or Nelson (2019) for discussion.

representation. To this end, let

$$\text{Rep} : \mathcal{P}(W) \times W \longrightarrow \mathcal{P}(\mathsf{Frml}_{\mathcal{L}_{\text{B}}})$$

be a function that selects the available representations of a proposition qua set of possible worlds relative to each world. This could, e.g., be a class of sentences that are intensionally isomorphic in the sense of Carnap (1947) or Lewis (1970). A formula of the form $\text{B}\varphi$ will then be true in a belief model M and world w, iff

$$\forall v(wRv \Longrightarrow M, v \vDash \varphi \,\&\, \varphi \in \text{Rep}_v(\|\varphi\|)).^{33}$$

On this view, $\text{B}\varphi$ and $\text{BT}t_\varphi$ will not always be semantically equivalent at w even though (TSW) will hold at every world, that is, $\|\varphi\| = \|Tt_\varphi\|$, because there is no guarantee that

$$\varphi \in \text{Rep}_v(\|\varphi\|) \Longleftrightarrow \text{T}t_\varphi \in \text{Rep}_v(\|\varphi\|).$$

In principle, such a semantics can individuate beliefs as finely as the sentences of the language but it can also allow for a much coarser individuation of beliefs. However, whereas the formal semantics thus settles the formal puzzle, that is, (OS) is no longer valid on such a semantics, there still remains the need for a principled philosophical explanation of why an agent may stand in the believing relation to ⟨Mary, Smartness⟩ without also standing in the believing relation to ⟨True, ⟨Mary, Smartness⟩⟩. Perhaps appealing to the way of believing may be useful to this effect and, in this case, the sketched semantics could potentially be combined with the semantics for one-level belief representation we outlined in Section 4.

7 Conclusion

The aim of the paper was to provide a more adequate semantics for belief and truth, that is, a semantics in which one can semantically distinguish between believing and believing-true. The main novelty of the semantics proposed in this paper was to explicitly appeal to the way of believing in the semantic evaluation of a formula of the form $\text{B}\varphi$ where the way of believing was extracted from the syntactic information provided by φ. We argued that this provides a neat semantic explanation of why believing and believing-true are not semantically equivalent. We take it that our semantics provides, at the very least, an interesting first step towards an adequate semantics for belief and truth.[34]

[33] Again, $\|\varphi\|$ is is the set of worlds in which φ is true, i.e.,

$$\|\varphi\| := \{w \in W \mid M, w \vDash \varphi\}$$

We can assume $\|\varphi\|$ to be defined at this stage of the semantic evaluation since φ is of lower complexity than $\text{B}\varphi$. Ultimately, the semantics would amount to a variant of Awareness semantics in the sense of Fagin et al. (1995) in which the Awareness set is constrained by a number of specific rules.

[34] It is perhaps worth noting that our proposal is not committed to the notion of full belief, as opposed to the notion partial belief or credences. The principal strategy underlying our semantics, that is, the idea of making the accessibility relation dependent on the way of believing can also be used to provide a semantics for truth and partial belief.

Acknowledgments

My research was supported by the ERC Starting Grant *Truth and Semantics* (TRUST, Grant n° 803684). I wish to thank Carlo Nicolai and Thomas Schindler for helpful comments on the paper, and Catrin Campbell-Moore, Ollie Tatton-Brown and the audiences of talks virtually given at Glasgow, Birmingham and Bristol for stimulating discussions of the material.

References

Asher, N. and Kamp, H. (1989). Self-reference, attitudes and paradox. In Chierchia, G., Partee, B. H., and Turner, R., editors, *Properties, Types, and Meaning. Vol. I: Foundational Issues*, pages 85–158. Kluwer.

Caie, M. (2012). Belief and Indeterminacy. *Philosophical Review*, 121(1):1–54.

Campbell-Moore, C. (2015). How to express self-referential probability. a Kripkean proposal. *The Review of Symbolic Logic*, 8(4):680–704.

Carnap, R. (1947). *Meaning and Necessity*. University of Chicago Press, Chicago.

Crimmins, M. and Perry, J. (1989). The prince and the phone booth: Reporting puzzling beliefs. *The Journal of Philosophy*, 86(12):685–711.

Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1995). *Reasoning about Knowledge*. MIT Press.

Field, H. (1994). Deflationist views of meaning and content. *Mind*, 103:249–285.

Field, H. (2006). Compositional principles vs. schematic reasoning. *The Monist*, 89(1):9–27.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.

Garson, J. W. (2001). Quantification in Modal Logic. In Gabbay, D. and Guenther, F., editors, *Handbook of Philosophical Logic*, volume 3, pages 267–323. Kluwer Academic Publishers. 2nd Edition. First published in 1984.

Gupta, A. and Belnap, N. (1993). *The revision theory of truth*. The MIT Press.

Halbach, V. and Leigh, G. (2020). *The Road to Paradox: A Guide to Syntax, Truth, and Modality*. Cambridge University Press. To Appear.

Halbach, V. and Welch, P. (2009). Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind*, 118:71–100.

Heck, R. K. (2020). Disquotationalism and the Compositional Principles. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*. Routledge. Forthcoming.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca and London.

Jerzak, E. (2019). Non-classical knowledge. *Philosophy and Phenomenological Research*, 98(1):190–220.

King, J. C. (2013). On fineness of grain. *Philosophical Studies*, 163(3):763–781.

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

Kripke, S. A. (1972). Naming and necessity. In Davidson, D. and Harman, G., editors, *Semantics of Natural Language*, pages 253–355. Springer Netherlands, Dordrecht.

Kripke, S. A. (1979). A puzzle about belief. In Margalit, A., editor, *Meaning and Use*, pages 239–283. Springer.

Künne, W. (2003). *Conceptions of Truth*. Oxford University Press.

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic*, 34:155–192.

Lewis, D. (1970). General semantics. *Synthese*, 22(1/2):18–67.

Nelson, M. (2019). Propositional attitude reports. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition.

Picollo, L. and Schindler, T. (2020). Is Deflationism Compatible with Compositional and Tarskian Truth Theories. In Nicolai, C. and Stern, J., editors, *Modes of Truth: The Unified Approach to Modality, Truth, and Paradox*. Routledge. Forthcoming.

Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, 15(1):47–87.

Stern, J. (2014a). Modality and Axiomatic Theories of Truth I: Friedman-Sheard. *The Review of Symbolic Logic*, 7(2):273–298.

Stern, J. (2014b). Modality and Axiomatic Theories of Truth II: Kripke-Feferman. *The Review of Symbolic Logic*, 7(2):299–318.

Stern, J. (2016). *Toward Predicate Approaches to Modality*, volume 44 of *Trends in Logic*. Springer, Switzerland.

Tarski, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research*, 4:341–376.

Visser, A. (1984). Semantics and the Liar Paradox. In Gabbay, D., editor, *Handbook of Philosophical Logic*, pages 617–706. Dordrecht.

Yablo, S. (1982). Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11:117–137.