Soames's argument 1 against strong two-dimensionalism

Robert Michels

Abstract This paper criticizes Soames's main argument against a variant of two-dimensionalism that he calls strong two-dimensionalism. The idea of Soames's argument is to show that the strong two-dimensionalist's semantics for belief ascriptions delivers wrong semantic verdicts about certain complex modal sentences that contain both such ascriptions and claims about the truth of the ascribed beliefs. A closer look at the formal semantics underlying strong two-dimensionalism reveals that there are two feasible ways of specifying the truth conditions for claims of the latter sort. Only one of the two yields the problematic semantic verdicts, so strong two-dimensionalists can avoid Soames's argument by settling for the other way.

Keywords Belief ascriptions · Propositional attitude ascriptions · Soames · Two-dimensionalism · Two-dimensional semantics

1 Introduction

Scott Soames's book *Reference and Description. The Case Against Two-Dimensionalism* delivers the most thorough critique of two-dimensionalism to date. This paper focuses on Soames's Argument 1, which targets one specific variety of the theory, strong two-dimensionalism.[1] Argument 1 is of central importance to Soames's critique of strong two-dimensionalism, as it provides the basic structure for his further arguments.[2] Soames coins the term strong two-dimensionalism to

---

[1] See Soames (2005b) for the original formulation of the argument. The argument is reproduced in Soames (2006) and, in a notational variant, in Soames (2005a).

[2] I agree with Dever concerning this point. See Dever (2007, pp. 9–10).

R. Michels (✉)
Département de philosophie, eidos, Université de Genève, Rue de Candolle 2,
1211 Genève 4, Switzerland
e-mail: mail@robert-michels.de

refer to a variant of two-dimensionalism that is based on ideas found in Chalmers (1996) and Jackson (1998). Argument 1 specifically targets the semantics for belief ascriptions associated with strong two-dimensionalism. This is quite surprising, since Soames acknowledges that an explicit semantics for belief ascriptions is put forward in neither of the two books.[3] Independently of whether Soames is correct in attributing strong two-dimensionalism to these two authors, the proposed strong two-dimensionalist semantics of belief ascriptions deserves serious attention, since it is the most straight-forward implementation of such a semantics available to a two-dimensionalist. Soames specifies this part of the semantics in one of a number of core theses that together constitute his characterization of strong two-dimensionalism. The main argument of this paper is based on the observation that these core theses fail to determine a unique semantic treatment of claims about the truth of a belief that is ascribed to an individual. As it turns out, strong two-dimensionalists are left with a choice of two alternative formal semantics for such claims. I will argue that only one of the two renders strong two-dimensionalists vulnerable to Argument 1 and that the other options allows them to avoid the argument.[4]

2 Two-dimensional semantics

The main focus of this paper is on the formal semantics underlying strong two-dimensionalism and on the claims made about such a semantics in the premises of Soames's Argument 1. As a consequence, many important philosophical questions about two-dimensionalism in general, e.g., the question of what it means that another world could have turned out as actual, will have to be left unanswered. The current section attempts to briefly introduce some fundamental concepts and philosophical motivations behind two-dimensionalism. As I see it, the two fundamental ideas of two-dimensional semantics are first, that each natural language expression is associated with two intensions, a *primary* and a *secondary intension*[5], and second, that these two kinds of intensions can be defined as functions that take two different kinds of possible worlds as their arguments. If the focus is on sentences rather then on sub-sentential expressions, these two basic ideas can alternatively be framed in terms of *primary* and *secondary propositions*.

The secondary intension corresponds to the familiar intension from possible worlds-semantics. Formally, it is represented by a function that assigns semantic values to possible worlds, which are called *worlds considered as counterfactual* or

---

[3] Soames finds Jackson to be '[…] less than fully explicit […]' (Soames 2005b, p. 172) about the semantics of attitude ascriptions, but argues that he should be understood as implicitly endorsing the strong two-dimensionalist proposal (See Soames (2005b, pp. 173–175)). Concerning Chalmers, Soames admits that '[…] he has very little explicitly to say about the semantics of propositional attitude ascriptions in *The Conscious Mind*.' (Soames 2005b, p. 235).

[4] A different perspective on Argument 1 has recently been offered in Dever (2007). I largely agree with Dever's diagnosis of his more abstract reconstruction of Argument 1, but I also think that Soames's original formulation of the argument raises important issues concerning the two-dimensionalist treatment of belief ascriptions that are lost in Dever's reconstruction.

[5] Some authors call them 1- and 2-Intensions or A- and B-Intensions, respectively.

simply *worlds*. In case of a rigid expression, this function assigns the same semantic value to all worlds of this kind. In this picture, which can be seen as a generalization of the theory of reference sketched in Kripke (1980), reference-fixing happens in the actual world. The value assigned to a rigid designator in the actual world hence determines the values assigned to it with respect to all other worlds. But according to two-dimensionalism, the actual world is merely one among several candidates that could have turned out to be the actual world. These candidate-worlds are called *worlds considered as actual*. Two-dimensionalists hold that different worlds considered as actual may fix different references for a given expression, so that the expression may have different secondary intensions with respect to different worlds considered as actual.

The primary intension of an expression is defined as a function from worlds considered as actual to semantic values. Its job is to assign to each expression the extension that it would have, had that world turned out as actual. Thereby, it also determines the secondary intension relative to that world considered as actual. There seems to be a consensus about the formal definition of the primary intension among recent proponents of two-dimensionalism, but there is a considerable controversy about its philosophical interpretation.[6]

One idea common to most variants of two-dimensionalism is that the primary intension captures an epistemic aspect of meaning and that secondary intensions capture an alethic, or metaphysical, aspect of meaning. This idea is also instrumental to the two-dimensionalist explanation of the necessary a posteriori: A sentence is necessary and a posteriori if and only if it has a contingent primary and a necessary secondary proposition. Similarly, a sentence is contingent and a priori if and only if it has a necessary primary and a contingent secondary proposition. This elegant semantic explanation is one of the biggest selling points of two-dimensionalism.

3 The formal apparatus

3.1 Syntax

The formal language used in this paper contains the following symbols:

A set of constants for atomic sentences $\mathsf{P} = \{p, q, r, \ldots\}$
A set of unary operators $\mathsf{U} = \{\neg, \Box, \diamond, @, B, Tr\}$
A set of binary operators $\mathsf{B} = \{\wedge, \vee, \rightarrow\}$
Parentheses (,)

Besides the usual logical symbols and the two modal operators $\Box$ for 'It is necessary that …' and $\diamond$ for 'It is possible that …', the formal language contains the 'actually'-operator @ and the two belief-related operators B and Tr.[7] The sentences of the language are recursively built up according to the following rules:

---

[6] See Chalmers (2006) for a comprehensive discussion of different interpretations of the primary intension and of Chalmers's favoured epistemic interpretation of two-dimensional semantics.

[7] Formulas that occur in prose sentences are sometimes mentioned, sometimes used. I will take the liberty of omitting the usual quotes.

Each $\phi \in \mathsf{P}$ is a sentence.

If ϕ is a sentence and $\bullet \in \mathsf{U}, \bullet(\phi)$ is a sentence.

If ϕ and ψ are sentences and $\circ \in \mathsf{B}, (\phi \circ \psi)$ is a sentence.

Parentheses will be omitted where permissible, as e.g., in $\neg\phi$.

3.2 Semantics

Sentences are evaluated with respect to a strong two-dimensionalist model $\mathfrak{M} = \langle W, C, v, b \rangle$. Each model contains a set of worlds considered as actual $c_1, \ldots, c_n \in C$ and a set of worlds considered as counterfactual $w_1, \ldots, w_n \in W$. There is an ongoing debate among two-dimensionalists about the relation between the two kinds of worlds. The members of C are usually taken to be centred worlds, ordered sets that contain an individual, a world and possibly other parameters that are needed for the evaluation of indexical or context sensitive expressions. The only indexical operator that appears in my discussion of Soames's Argument 1 is the 'actually'-operator @, which is sensitive to the world contained in the world considered as actual. For the purposes of this paper, all parameters except for the contained world can therefore safely be ignored. This means that in principle, we could do with models that contain only W, but I will nonetheless rely on models with both C and W, since the semantics is clearer if the two kinds of worlds are kept apart. An exception to this rule will be made for cases in which the relevant world is shifted to the world contained in the world considered as actual. The world considered as actual will in this case be taken to replace the relevant world, which allows for a simplification of the semantics for @. Lower case letters without subscripts, e.g., $w, w', \ldots,$ are used as variables ranging over the members of these sets. Each model contains a valuation function $v \in \{F : \mathsf{P} \mapsto \mathscr{P}(C \times W)\}$, which assigns to sentences in P a set of pairs of a world considered as actual and a world considered as counterfactual $\langle c, w \rangle$. The second function $b \in \{F : C \times W \mapsto \mathscr{P}(C)\}$ assigns sets of worlds considered as actual to a pair of a world considered as actual and a world. This function determines the primary propositions relevant to the evaluation of B. A word on the operator B and the presented model theory in general: The model theory used throughout this paper is tailor-made for the analysis of Soames's Argument 1. Since the sentences that need to be evaluated for this purpose involve only the beliefs of one individual, the model theory is not designed to account for the beliefs of multiple individuals. Throughout the paper, $B(\phi)$ is simply read as 'Mary believes that ϕ', since Mary is the only individual that figures in Soames's example sentences. The same goes for the operator Tr, the semantics of which will be introduced later.[8,9]

---

[8] The syntax and the semantics could easily be modified to account for the beliefs of multiple individuals by adding a set of indexed belief-operators to the language, by adding a set of individuals I to the model and by replacing b by $b' \in \{F : I \times C \times W \mapsto \mathscr{P}(C)\}$. Worlds considered as actual could then be identified with tupels of a world and an individual $\langle w, i \rangle$ (centered worlds).

[9] Note that models as defined above lack the accessibility-relation one comes to expect in a modal semantics. This simplification is unproblematic in the context of this paper, since such a relation plays no role in Soames's argument.

$$\mathfrak{M}, c, w \vDash p \quad \text{iff } \langle c, w \rangle \in v(p)$$

$$\mathfrak{M}, c, w \vDash \neg\phi \quad \text{iff } \mathfrak{M}, c, w \nvDash \phi$$

$$\mathfrak{M}, c, w \vDash (\phi \wedge \psi) \quad \text{iff } \mathfrak{M}, c, w \vDash \phi \text{ and } \mathfrak{M}, c, w \vDash \psi$$

$$\mathfrak{M}, c, w \vDash \Box\phi \quad \text{iff for all } w' : \mathfrak{M}, c, w' \vDash \phi$$

$$\mathfrak{M}, c, w \vDash @\phi \quad \text{iff } \mathfrak{M}, c, c \vDash \phi$$

$$\mathfrak{M}, c, w \vDash B(\phi) \quad \text{iff } b(\langle c, w \rangle) \subseteq [[\phi]]_1^{\mathfrak{M}}$$

As explained above, the world considered as actual is allowed to take the place of a world in the semantics for @.[10] The semantics for B captures the two-dimensionalist idea that we can say truly of someone that she believes ϕ, if she is adequately related to ϕ's primary proposition. The logical connectives $\vee, \rightarrow, \leftrightarrow$ and the possibility operator $\diamond$ are defined in the usual way. If it is clear from the context which model is being considered, the model will sometimes not be mentioned explicitly.

In the course of the paper, reference will be made to primary and secondary propositions, which correspond to the primary and secondary intensions of sentences. Based on the idea that propositions are sets of possible worlds, the primary intension $[[\phi]]_1^{\mathfrak{M}}$ and the secondary intension $[[\phi]]_2^{\mathfrak{M},c}$ of a sentence ϕ are defined in the following way:

$$[[\phi]]_1^{\mathfrak{M}} = \{c : \mathfrak{M}, c, c \vDash \phi\}$$

$$[[\phi]]_2^{\mathfrak{M},c} = \{w : \mathfrak{M}, c, w \vDash \phi\}$$

The primary proposition of a sentence ϕ relative to a model $\mathfrak{M}$ is the set of worlds considered as actual in which ϕ is true. It corresponds to the top-left to bottom-right diagonal of ϕ's *two-dimensional matrix* with respect to $\mathfrak{M}$ ($\mathfrak{M}$-*matrix* for short). The secondary proposition of a sentence ϕ with respect to a models $\mathfrak{M}$ and a world considered as actual c is the set of worlds considered as counterfactual, in which ϕ is true relative to c. This means that secondary propositions correspond to the rows of ϕ's $\mathfrak{M}$-matrix. The mentioned matrices are handy illustrations of the distribution of truth values in a two-dimensionalist model. Table 1 gives an example of such a matrix that illustrates a simple model for a necessary but a posteriori sentence q.

There are some significant terminological differences between the characterization of strong two-dimensionalism on pp. 133–139 of Soames (2005b) and the formal semantics presented here. Soames's characterization relies on Kaplan's concepts of *context* and *character* and, in his formulation of the semantics for belief ascriptions, on the concept of the primary intension. In the given semantics, the role

---

[10] The idea underlying the semantics for the 'actually'-operator @ is that 'actually' shifts the semantic focus to the current world considered as actual. A formally similar semantics for 'actually' is discussed in Cresswell (1990), chapter 3. The major alternative for a two-dimensionalist is a semantics that lets @ shift the world under consideration to a fixed world that is specified as a part of the model. This approach requires a different model theory, the classical exposition of which can be found in Davies and Humberstone (1980). The approach pursued in the current paper is briefly mentioned on p. 4–5 and in notes 4 and 5 on p. 26 of Davies and Humberstone (1980). My main argument can be made given either kind of model theory.

Table 1 $\mathfrak{M}_0$-matrix for q

	w_1	w_2
c_1	True	True
c_2	False	False

of contexts is played by the members of C. This should be fairly uncontroversial, since formally, Kaplanian contexts contain a world plus an individual, a position and a time. According to Soames, a sentence S's 'primary intension is a proposition which is true with respect to all and only those contexts C to which the Kaplan-style character of S assigns a proposition that is true at C.' (Soames (2005b), p. 133) Kaplan defines the character as the function which assigns a *content* to a sentence relative to a model and a context.[11] In the given semantics, the character of a sentence corresponds to a function which assigns secondary propositions to a sentence relative to a model and a world considered as actual. A sentence s can be said to be true in a world considered as actual in a model $\mathfrak{M}$ if and only if $\mathfrak{M}, c, c \vDash s$. Given this definition, the character of s assigns a secondary proposition that is true in a world considered as actual to exactly the worlds considered as actual that are members of s's primary proposition. Hence, the primary intension of a sentence directly corresponds to the primary proposition in the given semantics.

4 Soames's argument 1

Argument 1 is presented in five steps, the first four acting as premises, the last containing both the final premise and the conclusion. Since the first step merely reiterates the strong two-dimensionalist's semantics for belief ascriptions, I will focus on the remaining four steps:

Step 2   Since in every context C, the character of sentence (1a) expresses a truth with respect to C iff the character of sentence (1b) does too, the two primary intensions are identical, and the ascriptions
A believes that the actual husband of Stephanie Lewis was the actual author of Counterfactuals.
and
A believes that the husband of Stephanie Lewis was the author of Counterfactuals.
are necessarily equivalent. (In fact their secondary intensions, as well as their primary intensions, are identical.)

Step 3   Hence, the truth value of

    a.   It is a necessary truth that [if the actual husband of Stephanie Lewis was the actual author of *Counterfactuals* and Mary believes that the actual husband of Stephanie Lewis is the actual author of *Counterfactuals*, then Mary believes something true].
is the same as the truth value of

---

[11] Compare the definition of content on p. 546 and the definition of character on p. 548 of Kaplan (1989).

b. It is a necessary truth that [if the actual husband of Stephanie Lewis was the actual author of *Counterfactuals* and Mary believes that the husband of Stephanie Lewis is the author of *Counterfactuals*, then Mary believes something true].

   Since (b) is false, so is (a).

Step 4   Similarly, the truth value of

a. It is a necessary truth that [if Mary believes that the actual husband of Stephanie Lewis was the actual author of *Counterfactuals*, and if that belief is true, then the actual husband of Stephanie Lewis is the actual author of *Counterfactuals*].

   is the same as the truth value of

b. It is a necessary truth that [if Mary believes that the husband of Stephanie Lewis was the author of *Counterfactuals*, and if that belief is true, then the actual husband of Stephanie Lewis was the actual author of *Counterfactuals*].

   Since (b) is false, so is (a).

Step 5   Since, in fact the a-sentences in steps 3 and 4 are true, the strong two-dimensionalist theses T5a and T5b are not jointly true. (Soames (2005b), pp. 272–273. Underlines in Step 2 omitted.)[12]

In the terminology of the given formal semantics, the main claim of Step 2 is that, according to strong two-dimensionalism, the beliefs ascribed to A in the two sentences in italics have the same primary proposition and are therefore necessarily equivalent. Is this claim correct? Let p represent Soames's sentence (1a) 'The husband of Stephanie Lewis was the author of Counterfactuals.' The natural formalization of (1b) 'The actual husband of Stephanie Lewis was the actual author of Counterfactuals.' then is @ p.[13] In a model $\mathfrak{M}$, the two sentences have the same primary proposition. This is a direct consequence of the semantics for @ and the definition of the primary proposition. Since B is sensitive only to the primary proposition and p and @ p have the same primary proposition, $B(p)$ and $B(@\,p)$ are equivalent. They are also necessarily equivalent. The necessity of a sentence is determined by its secondary intension alone, which plays no part in determining the truth value of these two sentences.[14] The claim in parentheses is also correct. In any given model, $B\,(p)$ and $B\,(@\,p)$ have an identical primary proposition and are assigned the same secondary propositions.

Step 2 sets up the main part of the argument by drawing attention to the fact that the two belief ascriptions $B\,(p)$ and $B\,(@\,p)$ are necessarily equivalent in strong

---

[12] T5a and T5b are introduced on pp. 268–269 of Soames (2005b). The semantics for $\square$ and B captures the contents of the two definitions, so I will not quote them here.

[13] The formalization treats sentences of the form 'The actual F is the actual G' and 'Actually, the F is the G', where F and G are definite descriptions, as having the same meaning. This is unproblematic in the context of this paper.

[14] More precisely: Assume that logical truth is defined in the following way: ϕ is a logical truth iff for every $\mathfrak{M}$, for $c \in C \in \mathfrak{M}$, for $w \in W \in \mathfrak{M}$, $\mathfrak{M}, c, w \vDash \phi$. Given this definition, $(B(\phi) \leftrightarrow B(@\phi))$ is a logical truth of strong two-dimensionalism, since per definition $[[\phi]]_1^{\mathfrak{M}} = [[@\phi]]_1^{\mathfrak{M}}$.

two-dimensional semantics. The main aim of Argument 1 is to exploit this fact. Soames aims to show that strong two-dimensionalists commit themselves to an intuitively false semantic verdict about certain complex sentences containing belief ascriptions embedded under a necessity operator. In the following two sections, I propose a formalization of these complex sentences that makes them apt for semantic evaluation in strong two-dimensional semantics. The remaining steps of the argument will be discussed in Sect. 7.

5 The (a)- and (b)-sentences

In order to evaluate the claims made about the (a)- and (b)-sentences in Steps 3 and 4, the sentences have to be translated into the given formal language. To do this, I will introduce the new operator $Tr\,(\phi)$, which is read as 'Mary's belief that ϕ is true'. With Tr, we can formalize the four sentences from Steps 3 and 4 in the following way:

(3a) $\Box((@p \land B(@p)) \rightarrow Tr(@p))$
(3b) $\Box((@p \land B(p)) \rightarrow Tr(p))$
(4a) $\Box((B(@p) \land Tr(@p)) \rightarrow @p)$
(4b) $\Box((B(p) \land Tr(p)) \rightarrow @p)$

6 Truth conditions for *Tr*

While Soames specifies how a strong two-dimensionalist evaluates belief ascriptions, his characterization of the theory does not specify a semantics for *Tr*. The most likely explanation for this omission is that the strong two-dimensionalist semantics for B seems to imply that the truth or falsity of a subject's belief ϕ is determined solely by ϕ's primary proposition. But even if it is granted that the primary proposition does all the semantic work with respect to B, there is more to be said about the semantics of *Tr*. The point is easy to illustrate if one thinks of the two-dimensional matrix associated with a sentence that *Tr* is applied to. The cells of such a matrix correspond to world pairs of the form $\langle c, w \rangle$. The cells in the top-left to bottom-right diagonal represent those world pairs of this form for which $w \in c$. Since c already provides w for any such cell, these cells can simply be referred to as corresponding to c. I will call them *on-diagonal* cells. Since they represent the primary proposition of the sentence, there is no question to answer about the semantics for *Tr* relative to these cells of the matrix. *Off-diagonal* cells on the other hand represent world pairs $\langle c, w \rangle$ for which $w \notin c$. Since per assumption, the truth value of $Tr\,(\phi)$ is determined by ϕ's primary proposition alone, the semantics has to specify which on-diagonal cell is to be considered, if $Tr\,(\phi)$ is evaluated with respect to off-diagonal cells that do not correspond to the primary proposition. There are two *prima facie* plausible ways of doing this in the given semantics. Here is the first one:

PT1 $\mathfrak{M}, c, w \vDash Tr(\phi)$ iff $c \in [[\phi]]_1^{\mathfrak{M}}$ and $\exists c', w'$ so that $\mathfrak{M}, c', w' \models B(\phi)$.

The first part of the definiens of PT1 tells us that Tr is sensitive to the on-diagonal cell that corresponds to the provided world considered as actual c. The second part specifies the requirement that there be a pair of a world considered as actual c' and a world w' with respect to which the belief in question is held. Since it does not make sense to talk about the truth of a belief if that belief is not held at all, a requirement of this kind must be part of any semantics for Tr. The existential quantification is needed because there are cases in which we may want to assess the truth of a subject's belief with respect to a pair of a world considered as actual and a world, independently of whether the subject holds that belief with respect to this world pair. E.g., a strong two-dimensionalist might want to say that Johann's actually false belief that p could be true relative to a world pair where he holds the false belief that $\neg p$, even though p is true there.[15]

The second proposal is the following:

PT2 $\mathfrak{M}, c, w \vDash Tr(\phi)$
   iff for c' such that $w \in c', c' \in [[\phi]]_1^{\mathfrak{M}}$ and $\exists c'', w'$ so that $\mathfrak{M}, c'', w' \models B(\phi)$.

While PT1 tells us to look at the on-diagonal cell that corresponds to the current world considered as actual, PT2 tells us to look at the on-diagonal cell corresponding to the world considered as actual c' that contains the provided world considered as counterfactual.[16] PT1 and PT2 are the only sensible ways of defining truth conditions for Tr based on the primary proposition.

Strong two-dimensionalism does also allow us to specify a semantics based on the secondary proposition:

ST $\mathfrak{M}, c, w \vDash Tr(\phi)$ iff $w \in [[\phi]]_2^{\mathfrak{M},c}$ and $\exists c', w'$ so that $\mathfrak{M}, c', w' \models B(\phi)$

ST delivers the same results as PT1 and PT2 for world pairs that correspond to on-diagonal cells. With respect to pairs corresponding to off-diagonal cells, ST simply makes Tr sensitive to the world in the off-diagonal cell under consideration. This means that ST effectively makes Tr a vacuous operator, if there is a true instance of the corresponding belief ascription in the respective model: Given ST and a true instance of $B(\phi)$ in $\mathfrak{M}$, a world w is in $[[\phi]]_2^{\mathfrak{M},c}$ if and only if $\mathfrak{M}, c, w \vDash \phi$ if and only if $\mathfrak{M}, c, w \vDash Tr(\phi)$.

All three proposals deliver the same results with respect to world pairs representing on-diagonal cells. This means that only the different semantic treatment of world pairs representing off-diagonal cells can serve as a criterion for selecting among the proposals.

---

[15] A more compelling example can be given in a model that contains more than one individual: Johann's actually false belief that Johann does not exist could rightly be said to be true with respect to a world considered as actual that contains a subject other than Johann, in which Johann does not exist and in which he consequently entertains no beliefs.

[16] Disregarding the requirement that the relevant belief is held by the respective individual somewhere in the model, we can say that roughly, PT1 corresponds to Lewis's operator $\dagger$ and PT2 corresponds to Stalnaker's operator $\dagger$. See Stalnaker (1978), p. 319f and Lewis (1973), p. 63f.

I think that PT1 should be rejected, because it delivers intuitively wrong results with respect to world pairs representing off-diagonal cells: If it is true that Mary believes that giant pandas are an endangered species with respect to the world considered as actual c which contains the world w, where they are, would her belief also be true relative to a world considered as counterfactual w', where giant pandas are not endangered? Her belief should surely be false with respect to w', but PT1 tells us that her belief is true, since it is true in w. PT1's flaw is that it completely disregards the off-diagonal cells and hence the non-actual worlds considered as counterfactual.

PT2 and ST on the other hand provide two distinct, yet equally sensible ways of handling off-diagonal cells. PT2 gives Tr what I would like to call an empathetic reading. Applied to the given example, a PT2-based semantics tells us to think about whether Mary's belief would be true, if the world considered as actual c', which contains the non-actual world w' had turned out to be actual. If c' had turned out to be the actual world, her belief could not be said to be true, since then giant pandas would not actually be an endangered species. This reading of Tr asks us to consider what would be the case if the non-actual world that we are considering as counterfactual would have been actual.

ST tells us to keep considering c as the actual world and to think about whether Mary's belief corresponds to how things are in the counterfactual world w' from the perspective of c. Hence, ST also yields the expected result of rendering the belief false. The ST-based reading of Tr could be called factual.[17]

Which of the two remaining interpretations of Tr should a strong two-dimensionalist adopt? Since both appear to be coherent and since Soames's discussion of strong two-dimensionalist provides no clear incentive to prefer one over the other, I will from now on consider two versions of strong two-dimensionalism, one with PT2 and one with ST.

7 Analysing argument 1

Argument 1 aims to show that strong two-dimensionalism contradicts an intuitive semantic judgement of the (a)- and (b)-sentences in Steps 3 and 4.[18] I will only focus on Step 3 of the argument. This decision will be justified after a few considerations about the proposed analysis of the argument. Soames assumes that strong two-dimensionalists are committed to the following two claims about the (a)- and (b)-sentences in Step 3 and 4:

---

[17] The two proposals notably differ in their treatment of instances of @ in sentences of the form $Tr\,(@\;\phi)$. Given PT2, @ is vacuous if it occurs embedded under Tr, so the equivalence $Tr(\phi) \leftrightarrow Tr(@\phi)$ holds. This equivalence does not generally hold given ST.

[18] I am not convinced that we have reliable theory-independent intuitions about the truth values of complex modal sentences of the kind in question, but I accept Soames's claim for the sake of the argument.

1. (a) and (b) have the same truth value.
2. (b) is false and therefore (a) is also false.

In Step 5 of the argument, these semantic commitments are played out against the intuitive judgement that the (a)-sentences are in fact true, giving Argument 1 the form of a *reductio ad absurdum*: The idea is that assuming strong two-dimensionalism leads us to an intuitively unacceptable semantic verdict about the (a)-sentences. The conclusion, which is contained in Step 5, provides a diagnosis of what led to this wrong verdict: The combination of the strong two-dimensionalist core theses T5a and T5b[19], which specify the semantics for the necessity operator and for belief ascriptions respectively. In a notational variant of the argument published elsewhere, Soames draws the alternative conclusion that '[…] modal and epistemic operators in English do not take systematically different objects.' (Soames 2005a, p. 414.)

In this section, Soames's two claims are put to the test. In order to semantically evaluate the (a)- and (b)-sentences, we need an adequate model. Such a model should be constructed in a way so that it does not preclude the truth of the intuitive judgement that the (a)-sentences are true. Semantic intuitions supporting such a judgement have to be rooted in an intuitive theory of meaning. Given that such a theory should roughly track our true judgements about the world we live in, the model has to contain a world that accommodates the facts relevant to the semantic evaluation of the (a)-sentences as they are in our world. This will be the world w_1, in which the husband of Stephanie Lewis was the author of *Counterfactuals*, making p true in w_1.[20] The world considered as actual c_1 contains this world w_1. Any genuinely two-dimensionalist model has to provide more than one world that can be considered as actual. To keep things simple and since no further worlds are needed, the model will contain only one such world c_2. This world considered as actual contains the world w_2, in which the husband of Stephanie Lewis was not the author of *Counterfactuals* and in which p is therefore false. The sentences $B(\phi)$ and $Tr(\phi)$ will again be read as statements about beliefs held by Mary. In the (a)- and (b)-sentences, a belief of either p or $@p$ is ascribed to Mary. Each such ascription occurs in the antecedent of a conditional that is embedded under a necessity-operator. It follows from the truth-functional definition of $\rightarrow$ and the semantics for $\square$ that the corresponding (a)- or (b)-sentence is true if the respective belief is wrongly ascribed to Mary. In such a case, the whole argument collapses, since the second claim about strong two-dimensionalism in Step 3 (and Step 4) is falsified. The simplest way to exclude these cases and to save the argument from immediate collapse, is to let b be such that $B(p)$ and $B(@p)$ are true with respect to all world pairs. Since p and $@p$ share the same primary intension, this amounts to a single constraint on b. Based on these considerations, the model can be specified in the following way (see also Tables 2, 3):

---

[19] See footnote 12.

[20] Note that the truth value of p is not sensitive to the world considered as actual since it contains no indexical element.

Table 2 $\mathfrak{M}_1$-matrix for p

	w_1	w_2
c_1	True	False
c_2	True	False

Table 3 $\mathfrak{M}_1$-matrix for $@p$

	w_1	w_2
c_1	True	True
c_2	False	False

$$\mathfrak{M}_1 = \langle W, C, v, b \rangle$$
$$W = \{w_1, w_2\}$$
$$C = \{c_1, c_2\}$$
$$v(p) = \{\langle c_1, w_1 \rangle, \langle c_2, w_1 \rangle\}$$
$$b = \{\langle c_1, w_1 \rangle \mapsto \{c_1\}, \langle c_1, w_2 \rangle \mapsto \{c_1\}, \langle c_2, w_1 \rangle \mapsto \{c_1\}, \langle c_2, w_2 \rangle \mapsto \{c_1\}\},$$

Since the focus will be squarely on $\mathfrak{M}_1$ from now on, I will take advantage of a simplification of the (a)- and (b)-sentences allowed with respect to this model. $B(p)$ and $B(@\ p)$ are true with respect to all world pairs $\langle c, w \rangle$. Each such belief ascription appears in a conjunction in the antecedent of a conditional. This means that the truth value of each such conditional with respect to a world pair is completely determined by the other conjunct in the antecedent and by the consequent. With respect to $\mathfrak{M}_1$, the (a)- and (b)-sentences can therefore be simplified in the following way:

(3a*) $\Box(@p \rightarrow Tr(@p))$
(3b*) $\Box(@p \rightarrow Tr(p))$
(4a*) $\Box(Tr(@p) \rightarrow @p)$
(4b*) $\Box(Tr(p) \rightarrow @p)$

I will first state the truth conditions for these simplified sentences and then use $\mathfrak{M}_1$-matrices to illustrate their truth values with respect to the model. This will be done once for the version of strong two-dimensional semantics with PT2 and once for the version with ST. In both cases, the requirement that the belief is being held in at least one world pair in the relevant model, which is part of the truth conditions for both semantics for Tr, is trivially fulfilled with respect to $\mathfrak{M}_1$. Hence, this requirement can also safely be disregarded here. The purpose of this section is to investigate whether Soames's claims from Step 3 actually hold for the two versions of strong two-dimensional semantics and whether both verify the intuitive judgement about the truth of the (a)-sentences from Step 5. The focus will therefore be on the world considered as actual c_1, which represents our world considered as actual in $\mathfrak{M}_1$.

So why will only Step 3 of the argument be considered? The semantics for @ and the truth-functional definition of the material conditional $\rightarrow$ together guarantee the

Table 4 PT2: $\mathfrak{M}_1$-matrix for (3a*)

	w_1	w_2
c_1	False	False
c_2	True	True

Table 5 PT2: $\mathfrak{M}_1$-matrix for (3b*)

	w_1	w_2
c_1	False	False
c_2	True	True

truth of (4a) and (4b) with respect to any world considered as actual in which p is true. As p is true with respect to c_1, both sentences are necessarily true with respect to c_1, so Soames's second claim from Step 4 is false. This is the case with PT2 and with ST, which means that Step 4 is completely ineffective as a component of an argument against strong two-dimensional semantics as understood in this paper. Hence only Step 3 needs to be considered. A compact overview of the results of the following subsections can be found in Sect. 8.

7.1 Step 3: strong two-dimensional semantics with PT2

7.1.1 PT2: truth conditions for (3a*)

$\mathfrak{M}, c, w \vDash \Box(@p \rightarrow Tr(@p))$
   iff for all w', either $\mathfrak{M}, c, c \nvDash p$ or, for c' such that $w' \in c', c' \in [[@p]]_1^{\mathfrak{M}}$.

(See Table 4)

7.1.2 PT2: truth conditions for (3b*)

$\mathfrak{M}, c, w \vDash \Box(@p \rightarrow Tr(p))$ iff for all w', either $\mathfrak{M}, c, c \nvDash p$ or, for c' such that

$$w' \in c', c' \in [[p]]_1^{\mathfrak{M}}.$$

(See Table 5)

7.1.3 PT2: summary

Given PT2, (3a*) and (3b*) have the same truth values with respect to both c_1 and c_2. This means that Soames's first claim is true in $\mathfrak{M}_1$ concerning this variant of strong two-dimensional semantics. His second claim is true with respect to c_1, but false with respect to c_2. Given our focus on c_1, the second claim can also be considered true with respect to $\mathfrak{M}_1$.

Table 6 ST: $\mathfrak{M}_1$-matrix for (3a*)		w_1	w_2
	c_1	True	True
	c_2	True	True

Table 7 ST: $\mathfrak{M}_1$-matrix for (3b*)		w_1	w_2
	c_1	False	False
	c_2	True	True

7.2 Step 3: strong two-dimensional semantics with ST

7.2.1 ST: truth conditions for (3a*)

$\mathfrak{M}, c, w \vDash \Box(@p \to Tr(@p))$ iff for all w', either $\mathfrak{M}, c, c \nvDash p$ or $w' \in [[@p]]_2^{\mathfrak{M},c}$.

(Table 6)

7.2.2 ST: truth conditions for (3b*)

$\mathfrak{M}, c, w \vDash \Box(@p \to Tr(p))$ iff for all w', either $\mathfrak{M}, c, c \nvDash p$ or $w' \in [[p]]_2^{\mathfrak{M},c}$.

(Table 7)

7.2.3 ST: summary

The first claim from Step 3 does not hold in $\mathfrak{M}_1$: Given strong two-dimensional semantics with ST, (3a*) and (3b*) differ in truth value with respect to c_1. As a consequence, the second claim does also not hold, since (3a*) is true, (3b*) false with respect to c_1.

7.3 Step 5

As was noted earlier, Step 5 contains both the last premise of Argument 1 and its conclusion. The last premise claims that (3a) is true. This is the intuitive semantic judgement that strong two-dimensionalists are supposedly unable to account for. But the semantic evaluations show that strong two-dimensional semantics with ST can account for the intuitive judgement for (3a*) with respect to $\mathfrak{M}_1$. Strong two-dimensional semantics with PT2 on the other hand renders (3a*) false with respect to the same model. A strong two-dimensionalist who wants to accommodate the semantic intuition that the the (a)-sentences are true is therefore well advised to choose ST over PT2. Concerning the conclusion of the argument, a very similar point can be made. Given ST, the two crucial claims made in Step 3 are both rendered false in the model under consideration. Hence, as a strong two-dimensionalist, one should

	Step 3	(3a*)
with PT2	+/+	+
with ST	−/−	−

Table 8 Overview of the results

adopt ST rather than PT2, since given ST, Argument 1 has a false premise and is therefore unsound.

8 PT2 versus ST

Table 8 illustrates the results from the previous section: + indicates that Soames's claims apply to the version of strong two-dimensional semantics indicated on the left with respect to c_1 in $\mathfrak{M}_1$, − indicates that they do not. The two positions in the first column correspond to Soames's two claims from Step 3. The second column indicates whether Soames is right in claiming that the semantics fails to account for the intuitive truth of the (a)-sentence from Step 3.

There are three lesson to be drawn from the given analysis of Soames's Argument 1. First, if PT2 was the only feasible semantics for *Tr*, the argument (sans Step 4) would be effective. Second, strong two-dimensionalists who adopt ST can, contrary to Soames's claim, account for the semantic intuition that the (a)-sentences are true. Third, with ST in place, the strong two-dimensionalist's semantics fails to conform to the crucial assumptions made for the purpose of a reductio ad absurdum in Argument 1's premises, so the threat of the argument is completely neutralised. These seem good reasons to draw the conclusion that strong two-dimensionalists should adopt ST. The given argument however also offers a clear perspective to their adversaries: A convincing argument against ST would suffice to reinstate Argument 1. In the last section of this paper, I will discuss one important candidate for such an argument.

9 An argument against ST

Unlike PT2, ST semantically detaches *Tr* from the primary intension and hence from *B*. A strong two-dimensionalist who adopts ST is committed to the view that the question of whether one holds a certain belief is settled by one's relation to the relevant primary proposition, while the question of whether what one believes is true is settled by the relevant secondary proposition. This means that strong two-dimensional semantics with ST violates a certain metaphysical constraint on semantic theories, namely the constraint that in cases where a subject's belief is said to be true, there should be one unique proposition that plays both the role of the truth bearer and provides the object of the belief. As a consequence, a strong two-dimensionalist prima facie has to deny that a sentence like 'There is something such that Mary believes it and it is true.' can express a truth. This gives us the following argument against ST: ST violates the metaphysical constraint. The metaphysical

constraint must be met by any adequate semantic theory. Therefore, strong two-dimensionalists cannot adopt ST. I take this to be a valid argument. If the argument is also sound, the proposed strong two-dimensionalist reply to Argument 1 is blocked. In this section, I will briefly discuss three different responses to this argument against ST.[21]

The first response denies the first premise and attempts to explain away the conflict between the semantic theory and the metaphysical constraint. To do this, the quantifier in a sentence like 'There is something such that Mary believes it and it is true.' is given a non-objectual reading. This reading allows the friend of ST to account for the possible truth of this sentence without incurring ontological commitment to one unique proposition that is quantified over.[22] It is then claimed that this suffices to meet the metaphysical constraint. A general problem I see with this strategy is that the metaphysical constraint is, if at all, only met on the surface. The strategy does not ensure that the same entity is responsible both for the truth of Mary's having a belief and the truth of her belief, since it does not deal with such ontological issues at all.

The second strategy also denies the first premise. It relies on an equivalent reformulation of strong two-dimensional semantics in which both B and Tr operate on the same kind of semantic entity, the *compound proposition*. A compound proposition of a sentence ϕ can be defined as an ordered set $\langle [[\phi]]_1^{\mathfrak{M}}, \langle [[\phi]]_2^{\mathfrak{M},c_1}, \ldots, [[\phi]]_2^{\mathfrak{M},c_n} \rangle \rangle$, which contains both the primary proposition and the secondary propositions that ϕ is associated with in a model.[23] The model theory has to be modified accordingly, so that e.g., the valuation function assigns a compound proposition to each atomic sentence.[24] A reformulation along these lines moves the resulting semantics close to what Soames calls *hybrid two-dimensionalism*. Effectively changing strong two-dimensionalism to hybrid two-dimensionalism would beg the question against Soames, since Argument 1 is specifically aimed at

---

[21] From here on, strong two-dimensionalism will be taken to include ST, unless explicitly stated otherwise.

[22] Hofweber (2005) defends a form of ontologically non-committal quantification in the context of certain inferences, Rayo and Yablo (2001) more generally discuss non-nominal quantification and ontological commitment.

[23] Chalmers (2011) presents a two-dimensionalist theory of attitude ascriptions which employs complex propositions of a similar kind.

[24] Such a valuation function v' could be defined in the following way: $v' \in \{F : \mathsf{P} \mapsto \langle \mathscr{P}(C), \langle \langle c_1, \mathscr{P}(W) \rangle, \ldots, \langle c_n, \mathscr{P}(W) \rangle \rangle \rangle\}$. The first element in the sequence represents the primary proposition, the elements in the embedded sequence represent secondary propositions relative to the members of C. This gives us the following semantics:

$\mathfrak{M}, c, w \vDash p$ iff $w \in [[\phi]]_2^{\mathfrak{M}} \in \langle c, [[\phi]]_2^{\mathfrak{M}} \rangle \in v'(p)$

$\mathfrak{M}, c, w \vDash \neg \phi$ iff $\mathfrak{M}, c, w \nvDash \phi$

$\mathfrak{M}, c, w \vDash (\phi \wedge \psi)$ iff $\mathfrak{M}, c, w \vDash \phi$ and $\mathfrak{M}, c, w \vDash \psi$

$\mathfrak{M}, c, w \vDash \Box \phi$ iff for all $w' : \mathfrak{M}, c, w' \vDash \phi$

$\mathfrak{M}, c, w \vDash @\phi$ iff $\mathfrak{M}, c, c \vDash \phi$

$\mathfrak{M}, c, w \vDash B(\phi)$ iff $b(\langle c, w \rangle) \subseteq [[\phi]]_1^{\mathfrak{M}} \in v'(\phi)$

$\mathfrak{M}, c, w \vDash Tr(\phi)$ iff $\mathfrak{M}, c, w \vDash \phi$ and $\exists\ c',\ w'$ so that $\mathfrak{M}, c', w' \vDash B(\phi)$

strong two-dimensionalism. The proposed reformulation does however stay clear of this problem. The defining feature of hybrid two-dimensionalism is its semantics for B: A belief ascription $B(\phi)$ is true relative to a world pair c, w iff (a) the relevant subject accepts a sentence or mental representation M, such that $[[M]]_2^{\mathfrak{M},c} = [[\phi]]_2^{\mathfrak{M},c}$ and (b) $[[M]]_1^{\mathfrak{M}}$ and $[[\phi]]_1^{\mathfrak{M}}$ are appropriately related. [25] I take the basic idea behind this definition to be that both the primary and the relevant secondary proposition of a sentence contribute to the truth-conditions for B. This is not the case in the proposed reformulation of strong two-dimensional semantics. Secondary propositions are part of the compound proposition that B operates on, but they have no bearing on the operator's truth-conditions. Hence, the reformulation does not fall under Soames's definition of hybrid two-dimensionalism. This second strategy seems more promising than the first one: The reformulation of the semantics both incorporates ST and conforms to the metaphysical constraint by relying on compound propositions as the basic semantic entities.[26]

The third strategy is to deny the second premise and to consequently reject the metaphysical constraint. This move is less radical than it might seem. Strong two-dimensionalists have to reject a close relative of the constraint in any case, since their semantics for B and for $\Box$ operate on two different kinds of propositions. One could even argue that rejecting the metaphysical constraint allows them to tie their semantics for Tr closer to their semantics for $\Box$, thereby strengthening an intuitive link between necessity and the truth of beliefs. The crucial question about this approach is, which impact the denial of the constraint has concerning the adequacy of a semantic theory. This is a complex question that cannot be fully addressed here. In the following, I will only give a quick sketch of a general view of semantics that could be used to vindicate a strong two-dimensionalist's choice to deny the metaphysical constraint. On this view, the purpose of semantic theories is to provide precise models of certain aspects of meaning that adequately reflect our linguistic practice. In the standard case, the modelled aspect of meaning are the truth conditions for the sentences of our language. Plausibly, the most important criterion for the adequacy of a semantic theory of this kind is that the truth values assigned by it conform to our linguistic practice.[27] Whether a semantic theory makes use of one or multiple kinds of propositions in order to meet this criterion is then only of secondary interest. A semantic theory that makes use of multiple propositions might have the vice of not being ontologically parsimonious, but this theory-aesthetic disadvantage would carry no or only minimal weight concerning the overall adequacy of the theory. Unless it is shown that strong two-dimensional semantics with ST fails to deliver truth conditions that adequately reflect our linguistic practice, its proponents could hence justify rejecting the metaphysical constraint by adopting the sketched general view of semantics.

---

[25] See Soames (2005b), pp. 313ff for the original definition.

[26] Note that the second strategy could rightly be called ad hoc, if it turned out that the only motivation for the proposed reformulation was its utility for the sketched response to the argument against ST.

[27] Lewis could be read as advocating a view like this. See Lewis (1986), chapter 1.2, pp. 40–41.

References

Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.

Chalmers, D. J. (2006). The foundations of two-dimensional semantics. In M. García-Carpintero, & M. Josep (Eds.), *Two-dimensional semantics* (pp. 55–141). Oxford: Oxford University Press.

Chalmers, D. J. (2011). Propositions and attitude ascriptions: A Fregean account. Noûs (forthcoming). http://consc.net/papers/propositions.pdf. Accessed date 28.7.2010.

Cresswell, M. J. (1990). *Entities and indices, studies in linguistics and philosophy* (Vol. 41). Dordrecht: Kluwer Academic Publishers

Davies, M., & Humberstone, L. (1980). Two notions of necessity. *Philosophical Studies, 38*, 1–30

Dever, J. (2007). Low grade two-dimensionalism. *Philosophical Books, 48*, 1–16

Hofweber, T. (2005). A puzzle about ontology. *Noûs, 39*(2), 256–83

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Clarendon Press

Kaplan, D. (1989). Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishing

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell Publishing

Rayo, A., & Yablo, S. (2001). Nominalism through de-nominalization. *Noûs, 35*(1), 74–92

Soames, S. (2005a). Reference and description. In F. Jackson, & M. Smith (Eds.), *The Oxford handbook of contemporary philosophy* (pp. 397–426). Oxford: Oxford University Press

Soames, S. (2005b). *Reference and description: The case against two-dimensionalism*. Princeton: Princeton University Press

Soames, S. (2006). Kripke, the necessary aposteriori, and the two-dimensionalist heresy. In M. Garcia-Carpintero, & M. Josep (Eds.), *Two-dimensional semantics* (pp. 272–293). Oxford: Oxford University Press.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics, syntax and semantics* (Vol. 9, pp. 315–332). New York: Academic Press.