Maria Aloni*

On the Model-Theoretic Interpretation of a Mental State

DOI 10.1515/tl-2017-0002

How can we maintain that (i) Gregor Samsa turned into a beetle; (ii) imagine what it would be like if he turned into a horse; (iii) compare his intelligence with that of Sherlock Holmes while admitting that (iv) Gregor Samsa is a fictional character and therefore does not exist? Maier proposes a novel account of fictional statements within a "psychologistic" version of DRT (Kamp) which provides insightful answers to these questions while maintaining a uniform account of fictional and non-fictional names.

In Maier's proposal, statements are analysed as updates of (representations of) interpreters' mental states (so called ADSs). Interpreters' mental states can have different components, including anchors, belief components, imagination components and more. Ordinary statements typically update the belief component of such a state, fiction statements update the imagination component. Proper names are uniformly analysed as presupposition triggers in a DRT framework (Geurts). In DRT, presuppositions are treated as anaphors that need to be resolved or accommodated (van der Sandt): on Maier's proposal, fictional names differ from ordinary names in that they are resolved/accommodated within the imagination component of a mental state (locally) rather than to an internal anchor (globally). Maier's answer to the questions above rely on the fact that different components of a single mental state can share their discourse referents.

As an illustration, consider Maier's proposed analysis of the two sentences in example (28) here repeated as:

(1) a. Frodo is a hobbit born in the Shire.
b. Frodo is a fictional character invented by Tolkien.

In (2) we have a schematic representation of the ADS resulting from an update with these two sentences after resolution/accommodation of the proper names (see example (29) in Maier's article for the full analysis):

***Corresponding author: Maria Aloni,** Institute for Logic, Language & Computation (ILLC) and Department of Philosophy, University of Amsterdam, Science Park 107, 1098 XG Amsterdam, The Netherlands, E-mail: m.d.aloni@uva.nl

(2) $\{\langle_{\text{ANCH}}, [\ x \mid name(x, Tolkien);\ author(x)]\rangle,$
$\langle_{\text{IMG}}, [\ y, z \mid name(y, Frodo);\ name(z, Shire);\ldots]\rangle,$
$\langle_{\text{BEL}}, [\ fictional(y);\ invent(x, y)]\rangle, \ldots\}$

In (2), the presupposition triggered by the non-fictional name *Tolkien* is resolved to the (internal) anchor x, while the presupposition triggered by the fictional names *Frodo* and *Shire* are resolved within the imagination (IMG) component. All introduced discourse referents x, y and z are assumed to be in principle accessible from any subsequent component (in particular the third BEL-component).

Structures like (2) are the result of a sophisticated procedure of presupposition resolution and as such are a valuable source of information. On the other hand, ADSs are hardly explanatory without a model-theoretic interpretation. The first reason why the ADS in (2) needs a model-theoretical interpretation has to do with the free occurrence of x and y in the BEL-component. Somehow, we have to guarantee that this component is really about Tolkien as introduced in the ANCH-component and about Frodo as introduced in the IMG-component. For comparison, consider the possible logical rendering of the following discourse in dynamic semantics:

(3) a. A man is walking in the park. He is whistling.
b. $\exists x \phi(x) \wedge \psi(x)$

The formula in (3-b) cannot constitute a proper representation of the meaning of the discourse in (3-a) unless it is equipped with a model-theoretic semantics to the effect that the free variable in the second conjunct is dynamically bound by the existential quantifier in the first conjunct.

There is also a second reason why a model-theoretical interpretation is useful here and has to do with reasoning. We would like to be able to draw conclusions from representations like (2), for example that the subject in such a mental state believes that Tolkien is an author and that Tolkien invented Frodo. From ADSs as such, no conclusions can be drawn. Equipping ADSs with a model-theoretical interpretation would allow us to draw conclusions and reason about these structures.

Maier is well-aware of the importance of defining a model-theoretic interpretation for ADSs and indeed presents an explicit proposal in the appendix of the article. My first comment will focus on this part. My second (more general) point will address the relation between fiction, imagination and truth and the role of mental states in the interpretation of (fictional) statements.

1 ADS vs NBAS

Maier's mapping from an ADS to a corresponding model-theoretic object (a so-called NBAS) is a rather complex mechanism which roughly consists in mapping mental state components in an ADS to sets of dynamic information states (identified with sets of world-assignment pairs) in the NBAS in such a way that the assignment-parameter of an information state modelling a dependent mental component $\langle l, K\rangle$ must extend the assignment-parameter of one element of a state modelling a relevant background component (see Appendix for definitions, in particular definition (47)).

Reasoning is not explicitly discussed in the article, and it is potentially a non-trivial issue to arrive at a working definition of what conclusions can be drawn from a given ADS because of the complicating factor that a single ADS can contain different components with the same label. Nevertheless, by restricting attention to ADSs **K** which contain only one single belief component, $\langle$BEL, $K\rangle$, one plausible way to derive that an agent in such an ADS believes ϕ is by requiring that ϕ is entailed by all information states associated with K in all NBAS captured by **K**.

Now let us try to apply this definition to derive that an interpreter in the mental state represented by the ADS in (2) believes that (i) Tolkien is an author and that (ii) Tolkien invented Frodo. As far as I can see, neither of these two facts can be easily derived.

1. Let us start with the derivation of the belief that Tolkien is an author. Assuming the definition proposed above, this would be the case if the proposition that Tolkien is an author is entailed by all information states associated with the BEL-component in all NBAS captured by our ADS. As far as I can see this will not be obtained. The information that x is an author will be encoded in all information states associated to the ANCH-component, but nothing guarantees that it will be maintained in the states mapped to the BEL-component. This has to do with the way in which discourse referents are handled in this system. In standard dynamic semantics and DRT, discourse referents are inherently "descriptive," when interpreting the second conjunct in example (3) the pronoun *he* stands for *the man walking in the park*. Here instead any descriptive information about a discourse referent x is lost as soon as we move from one component to the other because while dependent components must (partially) share their assignment-part with some possibility in a background component, they are totally independent with respect to their world-part. This feature, which is possibly crucial for the solution Maier proposed for counterfictional imagination, appears to be problematic whenever we have an ADS **K** with the following

dependence structure: $\langle \text{BEL/ANCH}, K_1 \rangle \prec_K \ldots \prec_K \langle l, K_i \rangle \prec_K \ldots \prec_K \langle \text{BEL}, K_j \rangle$. It seems that any descriptive information encoded in $\langle \text{BEL/ANCH}, K_1 \rangle$ will be lost in $\langle \text{BEL}, K_j \rangle$.

2. Let us now turn to the derivation that the interpreter believes that Tolkien invented Frodo. Again assuming the definition proposed above, this would be the case if the proposition that Tolkien invented Frodo is entailed by all information states associated with the BEL-component in all NBASs captured by our ADS. As far as I can see, this again will not be obtained. All states associated with the BEL-component in our structure will encode the information that x invented y ($invent(x, y)$) (I am ignoring renaming of variables here), but it is hard to see what guarantees that in each single state both x and y get the right interpretation. The problem arises because x and y come from two different components which are referentially independent from each other and from the fact that information states "originate" from a single possibility in a single "background" state. Possibly, this problem is solved by the $\cup$-plus operator in Maier's definition (48) but the details are missing, so it is hard to verify. To illustrate, let Q_1, Q_2 and Q_3 be the information state functions in NBASs corresponding to the ANCH-, IMG- and BEL-components in our ADS (2), respectively. Let i_1 and i_2 be possibilities in $Q_1(i)$ and $Q_2(i)$ for some i in the background of both Q_1 and Q_2 in the relevant NBAS. Possibility i_1 will be a pair $\langle w_1, g_1 \rangle$ where w_1 is the actual world and g_1 an assignment function mapping x to an individual d_1 named Tolkien in w_1. Possibility i_2 will be a pair $\langle w_2, g_2 \rangle$ where w_2 is a possible world compatible with Tolkien's story-telling and g_2 an assignment function mapping y to an individual d_2 named Frodo in w_2. Both i_1 and i_2 are in the background for the interpretation of Q_3. The problem is that all possibilities in $Q_3(i_1)$ will assign d_1 to x, but no specific value for y is fixed in this state; and all possibilities in $Q_3(i_2)$ will assign d_2 to y, but no specific value for x is fixed in this state. But then our ADS is predicted to be compatible with a mental state in which the agent is in doubt on whether Tolkien invented say Gregor Samsa, and Frodo was invented by, say, Kafka.

2 Fiction, imagination and truth

In Maier's account, fiction statements are treated as "prescriptions to imagine", an idea which is technically implemented by analysing fiction statements as updates of the imagination component of an interpreter's mental state. Intuitively, however, the truth of a fiction statement like *Gregor Samsa turned into a beetle* does not seem to rely on the imaginative power of an interpreter but

rather on what is stated in Kafka's story. It is unclear how this framework can account for the intuitive distinction in truth value between (a) *Gregor Samsa turned into a beetle* and (b) *Gregor Samsa turned into a horse* since both these statements will be analysed as updates of the imagination component of an interpreter's mental state with no reference to Kafka's story-telling. Maier does not explicitly address the issue of truth in fiction, but in principle any analysis of fiction phenomena should be able to account for the difference in truth value between the sentences (a) and (b) above. I see two potential problems for Maier's approach with respect to this issue: (i) given the current representation of mental states it is unclear how to distinguish an imagination component based on some author's story-telling from an imagination component based on the agent's free imagination. In Maier's example (17), a book anchor x is introduced. Such anchors might be part of the solution here, but then we would have to make fiction-based IMG-components formally dependent on such anchors, possibly by having indexed IMG-components for fiction-based imagination vs non-indexed IMG-components for free imagination (this is not (yet) the case in example (17)); (ii) Even if we implement the suggestion above, in a psychologistic semantics, truth in fiction would have in any case to be a derived notion dependent on the content of interpreters' mental states, but is the mental state of an interpreter the relevant piece of reality when it comes to the evaluation of fiction?