Replies to the Critics*

Edward N. Zalta
Center for the Study of Language and Information
Stanford University

Menzel's Commentary

Menzel's commentary is a tightly focused, extended argument and it may be summarized as follows: (a) though Zalta gathers a range of phenomena under a small set of concepts, unfortunately, the framework is a possibilist one; (b) Zalta justifies possibilism by arguing that it provides the simplest and most natural explanation and analysis of such phenomena as ordinary modal discourse; but (c) by taking the modal operator as primitive, Zalta doesn't really offer any genuine analysis or explanation of modal discourse and so cannot establish the superiority of his possibilism.

With respect to (a), Menzel correctly points out that the theory's possibilism derives *not* from its commitment to abstract objects, but rather from its commitment to objects x that possibly exist but which don't in fact exist (i.e., to objects that satisfy the condition: $\Diamond E!x \,\&\, \neg E!x$). The theory doesn't explicitly assert that there are any of these objects, but when one adds ordinary modal intuitions such as 'There might have been something which is F' (i.e., $\Diamond\exists xFx$), the Barcan formula guarantees that there is something which might have been F (i.e., $\exists x\Diamond Fx$). If F is an existence-entailing property such as being a million-carat diamond or a talking donkey, it follows that there are things that might have existed (though which in fact don't exist).

While this shows that the theory is committed to possible objects, it is important to point out that there is a way to interpret the formalism that results in a theory that satisfies the demands of actualism! I sketched this interpretation on pp. 102–3 of the book, by suggesting that we read the predicate '$E!$' as 'being concrete' (instead of 'exists'), and that we read the quantifier '$\exists$' as 'there exists'. Under such an interpretation, the theory is committed neither to nonexistent objects nor unactualized possibles, even when we add ordinary modal intuitions. Recently, Bernard Linsky and I have investigated this interpretation in greater detail, and we've used it to defend the simplest quantified modal logic (a logic that includes the Barcan formulas). So there is some question about whether the formalism must be interpreted as possibilist, and I refer readers interested in the question to our forthcoming paper.[1]

With respect to (c). I think Menzel has a point. Given that I take the modal operator as primitive, there is clearly a sense in which I cannot analyze or explain intuitive uses of 'necessarily' in ordinary modal discourse. But there are other senses of 'analyze' and 'explain' for which it could be said that the theory does analyze and explain modal notions. For one thing, the primitive modal operator is governed by a set of axioms. By characterizing the modal operator's logical behavior, these axioms to some extent explain it. But more importantly, there is kind of analysis provided by the theorem that necessarily true propositions are true in all possible worlds. As Menzel points out, I define the notion of *possible world* in terms of the modal operator and define the notion of *truth at* a world in terms of encoding, and then derive certain theorems of world theory in the logic of encoding in a simple and direct manner. One such theorem establishes the equivalence of necessary truth (for propositions) and truth in all worlds. Its alternative form establishes the equivalence of possible truth and truth in some world. These equivalences allow us to move from Menzel's sentence ($1'$) to his sentence ($10'$). Of course Menzel is correct that ($10'$) contains numerous primitive modal operators. But surely there is some kind of analysis of modality that is offered by the equivalence of ($1'$) and ($10'$). ($1'$) after all is just about ordinary objects

*This paper is published in *Philosophical Studies*, **69**/2-3 (March 1993): 231-242. I would like to thank the members of the Program Committee for the 1992 Pacific Meetings of the APA for choosing to have an author-meets-critics session on my book. And of course I would like to thank each of the critics for the time they spent reading the book and the hard work they put into their commentaries. I have found the comments to be well-balanced and insightful. I have responded to the critics in the reverse order in which they delivered their comments at the meetings. This allows me to conclude my replies with a discussion about the relationship between metaphysical foundations and set theory, a topic raised by Tony Anderson.

[1] See "In Defense of the Simplest Quantified Modal Logic," forthcoming, *Philosophical Perspectives: Philosophy of Logic and Language*, J. Tomberlin (ed.), Atascadero: Ridgeview, 1993.

and their modal properties. But (10′) shows this to be equivalent to a complex web of statements asserting: that there are possible worlds, that they have modal properties of a certain sort, that they encode propositional properties of a certain sort, that there are objects at those worlds, that those objects exemplify certain properties at those worlds, etc. This may not be an analysis or explanation of modality in Lewis' sense, but it certainly does provide a much more fine-grained picture of modal reality than (1′) does. And that constitutes some kind of analysis I would think.

Moreover, this picture has virtues that actualist pictures of modal reality do not always have, namely, logical simplicity and extensionality. Its logical *simplicity* derives from the fact that it is based on the simplest S5 quantified modal logic with a single, fixed domain of objects. Formulas with existential quantifiers under the scope of modal operators have straightforward truth conditions, requiring that at other worlds there are objects that have there (in a definable sense) the relevant properties. The *extensionality* of the modal picture is exhibited in at least two ways. One way is that the definition of *possible world* carves out a domain of objects such that the necessary truth of propositions coincides with truth at all the objects in this domain (this stands in contrast to the actualist accounts offered in Fine [1977] and Menzel [1990]). The second way is that the identity conditions for the worlds themselves are extensionally defined: two worlds are identical iff they encode exactly the same propositional properties (this stands in contrast to the actualist accounts offered in Plantinga [1974] and Chisholm [1976], where it seems that exactly the same propositions may be true at distinct world-states p and $p \,\&\, (q \vee \neg q)$). These facts suggest that the kind of possibilism I advocate offers a simple and revealing way to systematically organize our intuitive modal discourse.

Deutsch's Commentary

Deutsch's commentary begins with a description of my treatment of propositions and attitudes and it concludes with a list of consequences that suggest the theory is 'unstable'. I think Deutsch's initial description doesn't reveal the underlying Fregean picture that drives my views on propositions and attitudes. Here is a semantic picture of a simple English sentence of the form 'Fa', where $\mathbf{d}(a)$ is the ordinary object that the subject term 'a' denotes, $\mathbf{s}(a)$ is the abstract object that serves as the

sense of 'a', $\mathbf{d}(F)$ is the ordinary property denoted by 'F' and $\mathbf{s}(a)$ is the abstract property that serves as the sense of 'F':[2]

$$\begin{array}{l}
\mathbf{PLUG}(\mathbf{s}(F),\mathbf{s}(a)) = \text{the sense of the whole sentence} \\
\qquad\quad\;\; \uparrow \quad \uparrow \\
\qquad\quad\;\; F \quad a \\
\qquad\quad\;\; \downarrow \quad \downarrow \\
\mathbf{PLUG}(\mathbf{d}(F),\mathbf{d}(a)) = \text{the denotation of the whole sentence}
\end{array}$$

I modify Frege's view by taking the denotation of the whole English sentence to be a structured entity, namely, a 0-place relation (a state of affairs or proposition), which has the denotation of the subject term *plugged* into the denotation of the predicate.[3] However, the *sense* of an English sentence is also a 0-place relation, which has the sense of the subject term plugged into the sense of the predicate. Consequently, the sense of the sentence is structurally identical to the denotation—only the constituents are different. So when most philosophers talk about "the proposition" that embodies "the semantic content" of the English sentence 'Fa', I follow Frege in distinguishing *two* kinds of semantic content—the proposition that is denoted and the proposition that serves as sense.[4] The former is the 'objectual' content of the sentence, the latter is the 'cognitive' content. The former is used in the *de re* readings of belief reports, the latter in the *de dicto* readings.

I do not employ these two kinds of semantic content when constructing the formal language used to represent English. Instead, the formal language behaves in a logically simple way—the terms and propositional sentences simply denote. There are terms and sentences in the formal language that respectively denote the objects and propositions denoted by English terms and sentences, and there are other terms and sentences in the formal language (involving underlines and subscripts) that respectively denote the abstract entities and propositions expressed by English terms and sentences. Thus, it is absolutely critical to separate the claims I make about the semantics of English from the claims I make about the

[2]The arrows pointing down signify the denotation relation, whereas those pointing up signify the sense relation.

[3]I also change Frege's picture by supposing that the sense of a term does not necessarily *determine* its denotation.

[4]This is somewhat of a simplification, since there are several ways to build up the proposition that serves as sense, depending on whether the terms are standing in *de re* or *de dicto* position. See [1988], pp. 170–3.

semantics of the formal language. For example, the definite descriptions of the formal language simply have a denotation (defined as that object, should there be one, that satisfies the description's matrix), whereas definite descriptions of English are analyzed as having both a sense and a denotation.

In the first part of his commentary, however, Deutsch seems to take remarks I make about the semantics of the formal language as remarks about the semantics of English. He says:

> Early on it is decided that not only names but also definite descriptions contribute only their denotations to the make up of propositions.

This is not quite accurate, and to see why, let us focus on the examples he cites:

(1) Socrates is wise.

(2) The son of Phaenarete is wise.

On my view, it is a matter of fact that (1) and (2) denote the same state of affairs or proposition (since 'the son of Phaenarete' denotes the same object as 'Socrates'). But these two sentences *express* different propositions because the senses of 'Socrates' and 'the son of Phaenarete' differ. So to say that descriptions "contribute only their denotations to the make up of propositions" is to ignore that English descriptions also contribute their senses to the proposition expressed.

Of course, if English sentences did have a single kind of semantic content, it would be wrong to identify the content of (1) and (2). Since Deutsch seems to presuppose that they do, he is led to claim that I have extended the thesis of direct reference to descriptions. He observes:

> Not even the most committed direct reference theorists have been so bold as to extend the thesis of direct reference to definite descriptions. . .

Yet I don't think I have extended the thesis of direct reference to definite descriptions. On my treatment, the reason a description has the denotation that it has (assuming it has a denotation) is: the denoted object *satisfies* the descriptional matrix. That is not a directly referential

treatment of descriptions.[5] One consequence of supposing that English sentences have two kinds of semantic content is that no 'appeal to illusion' is required to explain the puzzling behavior of 'Clemens is an author' and 'Twain is an author' in attitude contexts. The Fregean explanation is available: while these sentences denote the same state of affairs or proposition, different propositions serve as their senses. Thus, one could bear an attitude to the sense of one and not bear that attitude to the sense of the other.

Deutsch prefaces his list of the 'instabilities' in my account with the claim:

> The main source of instability, however, lies in Zalta's assumption that the cognitive stuff floating around a person's head can be made to play a semantic role.

But, strictly speaking, this is not accurate. I do not appeal to cognitive stuff in my treatment of the semantics of English, if by 'cognitive stuff' we mean the particular mental tokens that are involved in each person's cognitive apparatus. Rather, I appeal to what is better described as the *content* of this cognitive stuff. This content doesn't float around a person's head, but rather classifies and characterizes the stuff that does. I give a theory of this content, in terms of abstract objects, abstract properties, and the propositions that can be built out of such entities. There is a straightforward sense in which these entities represent, respectively, ordinary objects, ordinary properties, and propositions having ordinary objects and properties as constituents.[6]

[5] It looks like Deutsch and I may disagree about what the thesis of direct reference is. Deutsch identifies it as the thesis that a "name contributes only its referent as a propositional constituent", whereas I take it to be the negative theses that names do not have (Fregean) senses that determine their denotation, nor are their denotations determined by the satisfaction of some definite description for which they are an abbreviation. Even if we take Deutsch's formulation, however, I would still deny that I have extended the direct reference thesis to descriptions, since there are *two* propositions to consider, one to which the description contributes its denotation, and the other to which the description contributes its sense.

[6] Deutsch asks why *expressions* couldn't serve as the preferred vehicles of representation. But expressions, both types and tokens, are intrinsically meaningless entities– they have no intrinsic content. There is only a very weak sense in which they could represent something, namely, the sense in which a label can represent the thing it labels. That is not a very satisfying concept of representation. Abstract objects and properties, however, have an intrinsic content consisting of the properties they encode. By encoding properties they can 'direct' us toward things in the world that exemplify

Let me conclude with a brief discussion of two of the instabilities that Deutsch produces. The third instability he cites concerns the definition of *truly believes*, which is used to explain, for example, why persons K and L can 'have the same belief' that Sam Rhodes is a fine fellow even though the sense of 'Sam Rhodes' differs for both K and L. I represent their *de dicto* beliefs as relations to the propositions that serve as the sense of 'Sam Rhodes is a fine fellow'. K is related to a proposition having his sense of 'Sam Rhodes' as a constituent, whereas L is related to a proposition having his sense of 'Sam Rhodes' as a constituent. But the truth of both beliefs depend on the truth of the proposition *denoted* by 'Sam Rhodes is a fine fellow', which has Rhodes himself as a constituent. K and L 'have the same belief' in the loose, but definable, sense that the truth of their beliefs both depend on the truth of this one denoted proposition involving Rhodes himself. Deutsch points out that on this criterion, a person K who both believed (*de dicto*) that Twain is an author and (*de dicto*) that Clemens is an author (not knowing that they were the same person), would also have the same belief, contrary to intuition. But I don't see the problem here. K is related to two different 'sense' propositions—one containing his sense of 'Twain' and another containing his sense of 'Clemens'. But the truth of these beliefs depend on the proposition denoted by 'Twain (Clemens) is an author'. So there is a sense in which K has different beliefs, and a sense in which he has the same belief. Deutsch himself, as a direct reference theorist, defends the idea that if K believes Clemens is an author he just does believe that Twain is an author. All I have done is to define a sense of 'same belief' that the direct reference theorists could use when they want to say the belief that Clemens is an author is the same as the belief that Twain is an author.

The last instability Deutsch cites for my approach, which he says is the 'most serious and fundamental' one, involves the case of someone who uses language not with the intention to express a belief but who rather mocks such a use of the language. Deutsch suggests that my approach cannot distinguish this 'parroted' or mock speech from genuine speech. He bases this objection on the following claim:

> Zalta's doctrine's suggest a certain view of when one can truly attribute such and such a belief, assertion, or the like, to K. K may be said to have some belief about Samuel Clemens, for example, if K associates some (any) properties with 'Samuel Clemens' and the name maps to Samuel Clemens via causal and historical forces. It appears that Zalta would extend this doctrine to any words for things, even words such as 'snow' and 'white', for example.

those properties.

But I don't see how Deutsch derives this doctrine from my work. I have never tried to lay down the conditions that are *sufficient* for someone's having a belief. Rather, I have tried to explain the semantics of sentences that truly report beliefs. The analysis of such reports have left me with a theory about the *necessary* conditions of having a belief, but I have not supposed that these constitute sufficient conditions.[7] So I would deny that my work entails the view about the sufficient conditions for belief that Deutsch attributes to me. Some one who parrots words or engages in mock speech does not intend to express beliefs in the ordinary way, and I think that there is nothing in the views I have developed that force me to attribute beliefs to such people on the basis of such speech.

Anderson's Commentary

Anderson's commentary concludes with some very general issues about the relationship between set theory and metaphysical foundations. He begins, however, by describing problems that arise from the restrictions governing the property abstraction principle. These restrictions prevent the definition of new properties or propositions in terms of formulas $\phi(x)$ that have encoding predications and/or quantifiers binding relation variables. Focusing on these restrictions, Anderson: (a) offers 'new data' which seem to require the banished properties and propositions (he frames a story about encoding), (b) laments the loss of the rule of 'substitution for functional variables', and (c) notes that the defined identity symbol '=' cannot denote a relation, leading to the problem that:

> [if we are to be sure] that we have supplied an interpretation for an application of the logic, we must (somehow) know that any things which we take to be relations are *not*, however contingently, coextensive with identity.

[7] Maybe Deutsch is misled by the defnition of *truly believes*, which is stated in terms of necessary and sufficient conditions. One of the conditions required for S to truly believe ϕ is that S believe ϕ, but necessary and sufficient conditions for 'S believes ϕ' are not specified.

He further elaborates (c), citing the example '$x = y$ & p', where p is some sentence, saying:

> Does this form of expression correspond to a Zalta-relation? Well if p is false it does—the null relation—at least, it's coextensive with the null relation. But what if p is true? Well, no, because then it would be coextensive with identity. So the problem of determining in general whether or not an expression expresses (or is coextensive with) a Zalta-relation is as difficult a problem as is imaginable.

With regard to (a), I wonder what language couches the new data Anderson cites? The word 'encodes' doesn't occur in English with the technical sense the theory gives to it. So it sounds like Anderson's new data concerns the technical language of the theory, at which point his suggestion that such data be treated as of a higher 'type' or 'level' may be the best procedure to follow. With regard to (b), note that Anderson never explicitly formulates the rule of 'substitution of functional variables', and so I cannot say for sure whether or not it is preserved, and if not, how serious the loss is. Without such a formulation, we cannot determine which intuitive principle the rule is suppose to capture. If this rule is "a kind of second order analog of universal instantiation", then how does it differ from the rule of universal instantiation of terms that denote properties, relations and propositions, which is valid in the system? It is true that arbitrary expressions involving encoding formulas and quantified relation variables do not denote properties and so cannot be instantiated into a universal claims of the form '$\forall F\phi$'. But is this any different from the fact that arbitrary definite descriptions '$\imath y\psi$' do not denote objects, and so may not always be instantiated into universal claims of the form '$\forall x\phi$'?[8]

[8]Anderson also notes that the expression '$\exists G\exists x(Gx \,\&\, xF)$' does not designate a property of F. He then observes that the following instance of the abstraction principle for abstract objects,

$$\exists x[A!x \,\&\, \forall F(xF \equiv \exists G\exists x(Gx \,\&\, xF))],$$

cannot be *seen as derivable* by instantiating the 'third order' variable Φ in the following third order version of the abstraction principle for abstract objects:

TO2: $\forall\Phi\exists x[A!x \,\&\, \forall F(xF \equiv \Phi)]$

Here I would just deny that our evidence for the 'second order' abstraction principle (i.e., my formulation) is derived from the third order version of the principle (at least, I never explicitly appeal to the third order version when motivating the theory).

With regard to (c), I am not sure why Anderson thinks the problem of determining whether an expression expresses (or is is coextensive) with a Zalta-relation is 'as difficult a problem as is imaginable'. There is a clearcut procedure to follow whenever one wants to assert that an expression $\phi(x)$ with encoding encoding formulas, etc., denotes a relation, namely, add the assertion that there is such a relation as an axiom and prove the resulting theory consistent. Why is this as difficult a problem as is imaginable? Moreover, in the particular case Anderson describes, namely '$x = y$ & p', it is unclear why we would need such a relation. For the typical representation of the data, the primitive $=_E$ relation and the complex relation $[\lambda xy\, x =_E y \,\&\, p]$ seem to work just fine.[9]

Finally, I turn to the issues concerning the relationship between set theory and metaphysical foundations. Anderson produces a list of set-theoretic 'oddities' connected with the foundations, concluding with a proof that the addition of some basic set theoretical notions and principles to the foundations (which presently contain no such notions or principles) leads to a contradiction. Let me begin by discussing the 'oddities', all of which revolve around the theorem schema Anderson calls (CT) and its consequences (McM) and (UND). The theory generates these theorems because it is constrained to do so by Cantor's Theorem. If we momentarily allow ourselves the notions and techniques of set theory, we may intuitively describe the workings of the abstraction principle for abstract objects by saying that it generates and correlates a distinct abstract object with each distinct element in the power set of the set of properties. Now for some such distinct abstract objects a and b, the theory has to identify the properties $[\lambda y\, Rya]$ and $[\lambda y\, Ryb]$ (for arbitrary R), for otherwise there would be a one-to-one mapping from the power set of the set of properties into a subset of the set of properties, in violation of Cantor's Theorem. This is why the theory generates such theorems as (CT), (McM), and (UND).

[9]Anderson does cite the case of beliefs involving propositions in mathematics, such as that $2 + 2 = 4$, where such relations might be needed. But this involves the special case of mathematics, and the theory has something to say in this regard. Treating '$E!$' simply as a primitive, undefined property, the theory would take mathematical theories to be stories about certain entities, which 'exist' according to their respective stories. Then, numerals such as '2' and '4' denote characters of those stories, and propositions such as '$2 + 2 = 4$' could be formulated in terms of the relation $=_E$ of identity for existing objects. So I am not sure this example proves the need for a relation to anchor the more general, defined notion of identity.

Now, of course, from a point of view that presupposes that all objects must behave in accordance with Leibniz' Law (as it is expressed in terms of Russellian exemplification) and the laws of ZF, then one will find these results 'odd'. But the theory I have developed makes no such presupposition. It allows that there is the usual realm of objects that conform to Leibniz' Law and the laws of ZF, but it then goes on to axiomatize a realm of objects that fall outside the traditional conception of Leibniz and Russell. This realm provides a more general metaphysical foundation than set theory. From this point of view, the results Anderson cites simply become fascinating facts about the *limitations* of Leibniz's Law, Russellian exemplification, and set theory to make distinctions among the objects of a subtle, new kind. The logic of encoding and theory of abstract objects together sneak in between the cracks of Cantor's Theorem and generate distinct objects that can't be distinguished by exemplification predications and facts about set membership. So when Anderson calls the objects required by (UND) 'undistinguished' objects and says that they 'have *no* unique properties', one has to emphasize that the notion of 'having a property' is ambiguous in the context of the logic of encoding. Though these 'undistinguished objects' cannot be distinguished by properties that they exemplify, they can certainly be distinguished by the properties that they encode. So, from my point of view, on which the logic of exemplification and the theory of sets have no metaphysical priority over the logic of encoding and theory of abstract objects, certain abstract objects are not 'odd' but rather expose the limitations of standard techniques.

Anderson concludes by showing that if one adds a small portion of set theory to the metaphysical foundations I have proposed, one can derive a contradiction. His conclusion is:

> Maybe we have to choose between taking set theory seriously and taking Zalta's theory seriously. For my money, set theory has proved its mettle—and some of the combinatorial (e.g., the pigeon hole principle) and number-theoretic results that follow from it are, well, just true—and not just true in the set theory story.

To undermine this conclusion, let me distinguish set theory's *mathematical* mettle from its *metaphysical* mettle. No doubt set theory has mathematical mettle—it is so rich that we can define or model almost all other

mathematical theories from its perspective. But set theory's metaphysical mettle is extremely limited: (a) it does not offer a good theory of the exemplification relation,[10] (b) it does not offer even a good model of that relation (even with the help of possible worlds),[11] (c) it is not a good environment for constructing theories of propositions, states of affairs, situations, possible worlds, nonexistent objects, etc., and (d) it offers no special understanding into the nature of the other nonextensional contexts. Set theory has therefore yet to prove its metaphysical mettle.

Whereas banishing set theory from the foundations would be a drastic step for a lot of theories, it is not as drastic a step for the present one. In the present theory, sets may be identified as certain abstract objects, and set membership identified as a certan abstract relation. This results from treating mathematical theories as stories and mathematical objects and relations as characters of the story. I think none of the usefulness of mathematics is lost by treating ZF and other mathematical theories, such as Peano Number Theory, as stories and by treating sentences like '$\emptyset \in \{\emptyset\}$', '$2 + 2 = 4$', etc., as truths of the stories. However, the loss of the logic of encoding and the theory of abstract objects, were we to instead reject the present foundations in favor of ZF, would be much more grievous, since ZF doesn't have the metaphysical mettle that such a theory offers.

Bibliography

Chisholm, R., 1976, *Person and Object*, La Salle: Open Court

Fine, K., 1977, "Postscript" to Arthur Prior, *Worlds, Times, and Selves*, Amherst: University of Massachusetts

Menzel, C., 1990, "Actualism, Ontological Commitment, and Possible Worlds Semantics," *Synthese* **85**: 355–89

Plantinga, A., 1974, *The Nature of Necessity*, Oxford: Oxford University Press

Zalta, E., 1988, *Intensional Logic and the Metaphysics of Intentionality, Cambridge, MA: MIT Press/Bradford Books*

[10] To be a member of a set is not what it is to exemplify a property; no reason can be given as to why two distinct objects are members of the set of red things, for example, other than by saying that they both exemplify redness; etc.

[11] Set theoretic models of relations collapse necessarily equivalent relations; a relation is something that is 'predicable', 'unsaturated' or 'has gaps' or 'is repeatable', but none of these intuitions are captured by sets.