Seeing Through Opacity:
A Defense of the Russellian View of Propositional Attitudes

by

Leonard Jay Clapp

B.A., Philosophy
University of Wisconsin, Madison, 1988

Submitted to the Department of Linguistics and Philosophy
in Partial Fulfillment of the Requirements for the Degree of

Doctorate of Philosophy in Philosophy
at the

Massachusetts Institute of Technology
May 1994

Signature of author...

Certified by ....
Sylvain Bromberger, Professor of Philosophy
MIT Department of Linguistics and Philosophy
Thesis Supervisor

Certified by ....
Richard Cartwright, Professor of Philosophy
MIT Department of Linguistics and Philosophy
Thesis Advisor

Certified by ....
Robert Stalnaker, Professor of Philosophy
MIT Department of Linguistics and Philosophy
Thesis Advisor

Accepted by ....
George Boolos, Professor of Philosophy
MIT Department of Linguistics and Philosophy
Chairman, Committee on Graduate Students

Seeing Through Opacity:
A Defense of the Russellian View of Propositional Attitudes

by

Leonard Jay Clapp

Submitted to the Department of Linguistics and Philosophy
on February 11, 1994 in partial fulfillment of the requirements for the Degree of
Doctorate of Philosophy in Philosophy

Abstract

The primary purposes of my dissertation are, first, to motivate *Russellian theories* of propositional attitudes and propositional attitude ascriptions by criticizing *Fregean theories*, and second, to defend Russellian theories from *the arguments and problems posed by the phenomenon of opacity*. A theory of propositional attitudes and propositional attitude ascriptions is *Russellian* just in case it respects both the *Principle of Direct Reference*, and the *Principle of Semantic Innocence*. The Principle of Direct Reference states, roughly, that the sole contribution a referring term makes toward the *content expressed by* a sentence in which the term appears is the referent of the term, and not a Fregean *sense* or a "way of thinking of the referent of the term." And the Principle of Semantic Innocence states, roughly, that a term has the same referent or designation regardless of whether the term occurs inside or outside of the that-clause of an attitude ascription. Fregean theories, on the other hand, are theories based upon Frege's theory of sense and reference, and thus Fregean theories respect neither the Principle of Direct Reference, nor the Principle of Semantic Innocence. Fregean theories are often alleged to be superior to Russellian theories, however, on the grounds that Fregean theories can avoid the arguments and problems posed by the phenomenon of opacity.

In Chapter 1 I, first, motivate Russellian theories by reviewing Kripke's and Putnam's arguments in support of the Principle of Direct Reference and Davidson's arguments in support of the Principle of Semantic Innocence, and second I proffer a detailed explication of the arguments and problems posed for Russellian theories by *the phenomenon of opacity*, i.e. the phenomenon whereby a normal, understanding subject, say Odile, might assent to a sentence such as

(1) Twain is a great author.

yet dissent from a sentence such as

(2) Clemens is a great author.

The phenomenon of opacity poses three problems for Russellian theories: First, there are *epistemological arguments from opacity*; these arguments conclude that Russellian theories must be rejected because they lead to the contradictory result that Odile both holds and not hold the attitude of belief toward one and the same *content*. Second, there

is the problem of the *cognitive significance of occurrences;* i.e. how can a Russellian theory explain why Oscar assents to (1), yet dissents from (2)? And finally, third, there are *semantic arguments from opacity*; these arguments conclude that a Russellian theory cannot provide an adequate account of the truth conditions of attitude ascriptions such as 'Oscar believes that Clemens is a great author.'

In Chapter 2 I argue that *the Fregean strategy for avoiding the problems posed by the phenomenon of opacity* faces a fundamental difficulty. The Fregean strategy involves positing *cognitive values*--"senses," "modes of apprehension," or some such entities--to serve as *contents*--the objects of propositional attitudes--and as the *semantic values* of terms occurring inside the that-clauses of attitude ascriptions. Contrary to what is commonly supposed, however, theories which utilize the Fregean strategy are not immune to arguments based upon the phenomenon of opacity, and furthermore, in attempting to avoid these arguments, Fregean theories are forced to individuate cognitive values extremely finely. This is problematic for two reasons: *(i)* since cognitive values are the objects of propositional attitudes, the *publicity* of our beliefs, desires etc. is undermined; *(ii)* since cognitive values are also the semantic values of terms occurring the that-clauses of attitude ascriptions, the legitimacy of our ordinary propositional attitude ascribing practices is undermined. Given this fundamental difficulty, I suggest that Russellian theories, and the arguments which allegedly refute them, deserve careful reconsideration.

In Chapter 3 I examine the Russellian theories proposed by Salmon, Richard, and Crimmins and Perry. In attempting to avoid the problems posed by the phenomenon of opacity all of these theories invoke versions of the Fregean strategy. Each of these theories analyzes propositional attitudes as *ternary* relations which have subjects, Russellian propositions, and some kind of *mediator* as relata. The *mediators* posited by these theories play a role very similar to the role played by the Fregean theorists' *cognitive values*, and as a result each of these Russellian theories run afoul of a difficulty similar to the fundamental difficulty facing the Fregean strategy.

Finally, in Chapter 4 I illustrate how Russellian theories can avoid the arguments and problems posed by the phenomenon of opacity, yet also steer clear of the difficulties discussed in Chapter 3. I defend Russellian theories from *the epistemological arguments from opacity* by showing that these arguments are unsound. My rejection of these arguments relies heavily on Burge's externalist account of self-knowledge. Concerning *the problem of cognitive significance* I argue that, contrary to what is often claimed, Russellian theories are not worse off with regard to this problem than are Fregean theories. And lastly, I illustrate how a version of Kamp's "Discourse Representation Theory" can be utilized to defend Russellian theories from *the semantic arguments from opacity.*

Thesis Supervisor: Dr. Sylvain Bromberger, Professor of Philosophy

Acknowledgments

If the department of Philosophy at MIT were to give out an "MVP" (Most Valuable Philosopher) award, I would not be a candidate recipient, and rightly so. If, however, the department were to give out a "most improved" award, I suspect that I would be a candidate. I have many people to thank for this possible honor. First, I must thank Professor Sylvain Bromberger. Sylvain possesses the rare property of being both extremely intelligent, *and* extremely kind (λx (x is extremely intelligent & x is extremely kind)), and I am one of many who has benefited from Sylvain's instantiation of this rare property. It is not an overstatement to say that were it not for Sylvain's efforts, I would not have finished my degree at MIT. Second, I must thank Professor Richard Cartwright and Professor Robert Stalnaker. I fear that it is too easy to discern Professor Stalnaker's influence on this dissertation, while, ironically, I fear that it is too difficult to discern Professor Cartwright's. At any rate, their views and philosophical styles have had a considerable influence on my own. Third, I owe a special debt to Professor Terry Penner, who introduced me to Philosophy. I, at least, benefited from this introduction. Fourth, I must thank my parents and grandparents for their emotional and financial support. And finally, I thank all those who helped in more subtle, yet no less important, ways. I hope and trust that at least some of you know who you are.

Table of Contents

Dear Colleague,

... Concerning sense and denotation, I see nothing but difficulties which I cannot overcome. ... I believe that in spite of all its snow fields Mount Blanc itself is a component part of what is actually asserted [by the sentence] 'Mount Blanc is more than 4,000 metres high'. We do not assert the thought, for this is a private psychological matter: we assert the object of the thought, and this is, to my mind, a certain complex (an objective proposition, one might say) in which Mount Blanc is itself a component part. If we do not admit this, then we get the conclusion that we know nothing at all about Mount Blanc. This is why for me the denotation of a [sentence] is not the true, but a certain complex which (in the given case) is true. In the case of a simple proper name like 'Socrates', I cannot distinguish between sense and denotation; I see only the idea, which is psychological, and the object. Or better: I do not admit the sense at all, but only the idea and the denotation. . .

Yours sincerely
Bertrand Russell[1]

Introduction: The Russellian and Fregean Approaches to Propositional Attitudes and Propositional Attitude Ascriptions.

The Russellian approach to propositional attitudes and propositional attitude ascriptions is enjoying a resurgence. Writers such as Kaplan, Salmon, Soames, Fodor, Richard, Perry and Crimmins have espoused theories at the core of which is the thesis that *Russellian propositions* are the semantic values of declarative sentences and the objects of propositional attitudes. (Russellian propositions are propositions which may contain physical objects as constituents.[2]) The primary virtue of the Russellian approach--and the reason for its resurgence--is that it respects both the *Principle of Direct Reference* and the *Principle of Semantic Innocence*. The Principle of Direct Reference states, roughly, that the sole contribution a referring term makes toward the *content expressed by* a sentence in which the term appears is not a Fregean *sense* or a "way of thinking of the referent of the term," but instead simply is the referent of the term. (The *content expressed by* an occurrence of declarative sentence is what one who believes what is expressed by the sentence holds the attitude of belief toward.) And the Principle of Semantic Innocence states, roughly, that a term has the same referent or designation regardless of whether the term occurs inside or outside the that-clause of an attitude ascription. Russellian theorists maintain, correctly on my view, that an adequate account of propositional attitudes and propositional attitude ascriptions must respect these two principles.[3]

It is often alleged, however, that Russellian theories, theories which respect the principles of direct reference and semantic innocence, can be refuted by arguments based upon the phenomenon of opacity. For example, suppose that a normal understanding subject, Odile, assents to an occurrence of

(1) Twain is a great author.

yet dissents from an occurrence of

(2) Clemens is a great author.

It follows from Odile's behavior that

(3) Odile believes that Twain is a great author.

is true, while

(4) Odile believes that Clemens is a great author.

is false. It seems that a Russellian theory, however, is forced to maintain that (3) and (4) have the same truth conditions: It follows from the Principle of Direct Reference (and a standard principle of compositionality) that (1) and (2) express the same Russellian proposition--they express the same *content*. Consequently, it follows from the Principle of Semantic Innocence that the that-clauses of (3) and (4) express the same Russellian proposition. Therefore (again assuming a standard principle of compositionality) (3) and (4) must present the same Russellian proposition, and thus they must have the same truth conditions. These results are deemed to be unacceptable. If Odile assents to an occurrence of (1), yet dissents from an occurrence of (2), then these occurrences must express distinct contents, and furthermore it follows from Odile's behavior that she believes that Twain is a great author, but does not believe that Clemens is a great author. Consequently, the argument concludes, the Russellian approach to propositional attitudes and propositional attitude ascriptions must be rejected.

Arguments such as the above serve as the primary motivation for the Fregean approach to propositional attitudes and propositional attitude ascriptions; Fregean theories--theories which are based upon Frege's theory of sense and reference--are

specifically designed to solve the problems and preclude the arguments which are posed by the phenomenon of opacity. Fregean theories maintain that the objects of propositional attitudes are not Russellian propositions composed of referents and properties, but rather are (something like) Fregean *thoughts*, which are composed of (something like) Fregean *senses*. Thus, contrary to the Principle of Direct Reference, under the Fregean approach the *content* expressed by an occurrence of a referring term is not the referent of the term, but rather is the sense expressed by the term, where a sense is something like a "mode of presentation," or a "way of thinking of the referent." Assuming that 'Twain' and 'Clemens' express distinct senses, the Fregean approach is thus able to distinguish the contents, or thoughts, expressed by (1) and (2). Furthermore, contrary to the Principle of Semantic Innocence, under the Fregean approach a referring term has distinct referents depending upon whether the term occurs inside or outside the that-clause of an attitude ascription; under the Fregean approach a term occurring inside the that-clause of an attitude ascription does not refer to its "ordinary referent," but instead refers to its sense, or its "oblique referent." Because the Fregean approach violates the Principle of Semantic Innocence in this way, Fregean theories are seemingly able to distinguish the truth conditions of (3) and (4): Under the Fregean approach (3) states that Odile holds the attitude of belief toward a thought which contains as a constituent one "way of thinking" of Twain, while (4) states that Odile holds the attitude of belief toward a distinct thought which contains as a constituent another "way of thinking" of Twain. Thus the Fregean approach is seemingly able to solve the problems and preclude the arguments posed by the phenomenon of opacity, but only at the cost of the principles of direct reference and semantic innocence.

The primary purposes of this dissertation are, first, to motivate the Russellian approach to propositional attitudes and propositional attitude ascriptions by criticizing the Fregean approach, and second, to sketch a Russellian Theory which avoids the arguments and problems posed by the phenomenon of opacity. The dissertation proceeds as follows.

In Chapter 1 I first motivate the Russellian approach by reviewing some of the arguments in support of the principles of direct reference and semantic innocence, and then I explicate the arguments and problems posed by the phenomenon of opacity in some detail. In Chapter 2 I further motivate the Russellian approach by showing that what I call *the Fregean strategy for precluding the arguments from opacity* faces a fundamental difficulty. That is, in Chapter 2 I set aside the Russellian arguments in support of the principles of direct reference and semantic innocence, and argue that there is an independent reason for rejecting the Fregean approach; even if one ignores the fact that Fregean theories come at the cost of the principles of direct reference and semantic innocence, there is a compelling reason for rejecting such theories. In Chapter 3 I turn to the Russellian approach and examine the Russellian theories recently proposed by Salmon, Richard, and Crimmins and Perry. Each of these Russellian theories reverts to employing aspects of the Fregean strategy in attempting to avoid the arguments and problems posed by the phenomenon of opacity, and as a result each of the these theories encounters the difficulty with the Fregean approach discussed in Chapter 2. Finally, in Chapter 4 I sketch a Russellian theory which avoids the arguments and problems posed by the phenomenon of opacity, and yet also steers clear of the pitfalls encountered by the Fregean strategy.

1.0 Direct Reference, Semantic Innocence, and the Arguments and Problems Posed by the Phenomenon of Opacity.

In this chapter I will discuss what I take to be the central arguments for and against the Russellian approach to propositional attitudes and propositional attitude ascriptions. In sections 1.1 and 1.2 I motivate the Russellian approach by reviewing some of the arguments in support of the principles of semantic innocence and direct reference. In section 1.3 I introduce what I call *The Naive Russellian Theory*, and I explicate the

arguments and problems posed by the phenomenon of opacity in terms of this Russellian theory.

1.1 The Principle of Direct Reference.

The notion of Direct Reference is due to Kaplan. Kaplan explains that a term is *directly referential* just in case "the relation between the [term] and the referent is not mediated by the corresponding propositional component, the content of what is said. . . . The directly referential term goes *directly* to its referent, directly in that it does not first pass through the proposition."[4] Note that to say that a term is directly referential is not to claim that there is *nothing* which determines, or *fixes*, the referent of the term[5]; indeed the Principle of Direct Reference is closely associated and often confused with just such a view; viz. the causal theory (or picture) of reference.[6] Rather to say of a term that it is directly referential is to say that the *content* expressed by an occurrence of the term *is* its referent and not whatever it is that determines its referent. (Again, I use 'content' as a theory neutral term designating the objects, or relata, of propositional attitudes and--what I take to be the same thing--that which is *expressed by* occurrences of declarative sentences. Hence Fregean thoughts and Russellian propositions are two kinds of *content*. Moreover, the content expressed by an occurrence of a *term* is the contribution the term makes toward the content expressed by an occurrence of a sentence containing the term.)

If there are directly referential terms, these terms serve as counterexamples to the Fregean conception of reference and content. According to this conception, the content expressed by an occurrence of a referring term is a *sense* where, again, a sense is a "way of thinking" of the referent, or a "mode of presentation" of the referent, etc. I will call such "ways of thinking" *cognitive values*.[7] (Thus as I use 'cognitive value', senses are a kind of cognitive value.) Under the Fregean conception of reference and content, the cognitive value of a referring term serves two primary functions: (i) it determines the

referent of the term (if the term has a referent), and (ii) it serves as the content expressed by occurrences of the term (when the term does not occur within the that-clause of a propositional attitude ascription). Theories which claim that some terms are directly referential, however, deny that all occurrences of referring terms express cognitive values which serve these two functions; rather such theories maintain that the content expressed by an occurrence of a directly referring term is the referent of the occurrence (if it has one), and not whatever it is that determines the referent. Thus theories which maintain that some terms are directly referential are incompatible with the Fregean conception of reference and content.

Russellian theories maintain not only that some referring terms are directly referential, but that all referring terms are directly referential.[8] That is, Russellian theories maintain the following principle:

> The Principle of Direct Reference: The content expressed by an
> occurrence of a proper name, a deictic pronoun, or a demonstrative is the
> referent of the occurrence, if it has a referent.[9]

(For the sake of simplicity I will focus on cases involving proper names, though I intend the views proposed here to be applied to all referring terms.) The rejection of the Fregean conception of reference in favor of views which respect the above principle has been to a large degree motivated by, first, Kripke's arguments in *Naming and Necessity*, and second Putnam's famous thought experiment in "The Meaning of 'Meaning'."[10] Both of these works provide convincing arguments against the Fregean conception of reference and content by attacking the Fregean dictum that "sense determines referent," or more generally that "cognitive value determines referent"; these works present arguments showing that the Fregean conception is incorrect in identifying the referent determining mechanism of an occurrence of a term with the content expressed by the occurrence. As the Principle of Direct Reference represents the most plausible alternative to the Fregean

conception of reference and content, these arguments against the Fregean conception constitute compelling arguments in favor of the Principle of Direct Reference.

Kripke's arguments are directed against "descriptivist" theories of reference and content for proper names. According to such theories, proper names function as "abbreviations" for definite descriptions. Thus, according to a descriptivist, cognitive values are certain descriptions (where *descriptions* are the contents expressed by definite descriptions); i.e. a subject's "way of thinking" of the referent of a term (when he uses the term) is the content of a certain definite description which denotes the referent of the term. (Frege certainly encouraged this identification, at least in the case of proper names.[11]) Thus according to the descriptivist, referring terms present descriptions, and these descriptions serve both of the functions called for by the Fregean conception of reference and content: descriptions serve as (i) that which determines the referent of referring terms (the referent is the object denoted by the description presented by the term), and (ii) the content expressed by occurrences of referring terms. Kripke argues, however, that such descriptions can serve neither of these functions: descriptions cannot adequately determine the referents of occurrences of referring terms, nor can they serve as the contents expressed by occurrences of the terms.[12]

A description cannot serve to determine the referent of an occurrence of a referring term because a subject can use a term to refer to an entity even if he does not, in using the term, utilize a description which denotes the entity--even if he does not think of the entity *via* a description which denotes the entity. And, conversely, a subject might fail to refer to an entity by using a term even if he does, in using the term, utilize a description which denotes the entity--even if he does think of the entity via a description which denotes the entity. For example, a subject might, in using 'Gödel', refer to Gödel, even though (almost) everything he thinks about Gödel is false. Suppose that our subject utilizes the description *the guy who discovered the incompleteness of arithmetic* in using the name 'Gödel'--he thinks of the referent of 'Gödel' as the guy who discovered the incompleteness

of arithmetic. Further suppose that, contrary to what we all assume, Gödel had not actually discovered the incompleteness proof, but rather had copied the proof from Schmidt. In this situation the description *the guy who discovered the incompleteness of arithmetic* would not denote Gödel. Nonetheless, if our subject were to utter the sentence 'Gödel was a logician', he would be talking about, and referring to, Gödel, not Schmidt; in uttering this sentence, our subject would refer to Gödel, even though he would not, in using 'Gödel', utilize a description which denotes Gödel. Therefore a subject's utilizing a description which denotes an entity is not *necessary* for using a term to refer to the entity.

The same example illustrates that the converse is also true; i.e. in using a referring term a subject's utilizing a description which denotes an entity is not *sufficient* for the subject's using the term to refer to the denoted entity. In the above example, the description *the guy who discovered the incompleteness of arithmetic* denotes Schmidt, yet despite this our subject would refer to Gödel, not Schmidt, were he to utter an occurrence of 'Gödel was a logician.' Examples such as this illustrate that the descriptions utilized by subjects in using referring terms cannot serve as the mechanism which determines the referents of those terms.

Kripke also argues that the content of a definite description cannot serve as the content expressed by an occurrence of a referring term; i.e. where referring term ρ allegedly "abbreviates" definite description δ, the content of δ cannot serve as the content of ρ. Kripke, borrowing from Frege's example concerning 'Aristotle', writes that

> if 'Aristotle' meant *the man who taught Alexander the Great*, then saying 'Aristotle was a teacher of Alexander the Great' would be a mere tautology. But surely it isn't; it expresses the fact that Aristotle taught Alexander the Great, something we could discover to be false. So, being the teacher of Alexander the great cannot be part of [the sense of] the name.[13]

Similarly, if the content of 'Aristotle' were the description the man who taught Alexander the Great, an occurrence of 'Aristotle did not teach Alexander the Great' would express

some kind of contradiction; an understanding subject would know a priori of an occurrence of this sentence that it was false. But an understanding subject could not know *a priori* of an occurrence of this sentence that it was false. Thus Kripke's arguments show that the content of a definite description which a referring term allegedly "abbreviates" can serve neither as the mechanism determining the referent of the term, nor as the content of the term. This conclusion provides support for the Principle of Direct Reference because such descriptions are the most plausible candidates for the role of cognitive values. What else could plausibly both determine the referent of the term and also serve as the content presented by the term? How else are we to understand the oft repeated, but seldom explained, claim that the sense, or a cognitive value, presented by a term is a "way of thinking" of the referent of the term?

Putnam's objection to the Fregean conception of reference and content is more broad in scope than Kripke's objection. Putnam's thought experiment shows that the *narrow*, or internal, psychological state of a speaker does not determine the referent of a term uttered by the speaker; two subject's in the same narrow psychological state can refer to distinct entities.[14] Consequently, if--as surely must be the case--the cognitive value "grasped" by a speaker must be somehow represented psychologically (i.e. if two subjects cannot grasp distinct cognitive values while being in the same narrow psychological state), then cognitive value does not determine referent. Putnam initially employed his thought experiment to show that the narrow psychological state of a subject who understands and uses a natural kind term does not determine the referent of the term, but the experiment can be modified to apply to any kind of referring term: Imagine that somewhere in the universe there is a planet, Twin-Earth, which is a molecule for molecule match of Earth. Now imagine that both an Earthling and his Twin-Earth *doppelgänger* utter 'Clinton has won.' By assumption, both subjects are in identical narrow psychological states--their brains are molecule for molecule matches with molecule for molecule matching histories. Nonetheless the subject on Earth refers to Clinton with his

utterance of 'Clinton', and the subject on Twin-Earth refers to Clinton's Twin-Earth *doppelgänger* with his utterance of 'Clinton'. Thus despite being in identical states, the two subjects refer to different entities. Consequently, if it is assumed that the narrow psychological state of a subject determines the cognitive value grasped by a subject, it follows that the Fregean conception of reference and content must be rejected.

Putnam's argument assumes that narrow psychological state determines "grasped" cognitive value--no two subjects in the same narrow psychological state can be grasping distinct cognitive values. If this assumption is rejected, then Putnam's argument against the Fregean conception of reference and content is undermined. Of course if cognitive values are identified with some kind of narrow psychological states, e.g. "cognitive roles," then Putnam's assumption is trivial: if cognitive values are identified with psychological states, then it is of course true that two subjects in the same narrow psychological state must be grasping (or utilizing) the same cognitive value. But if cognitive values are identified with Platonic entities, such as Fregean senses, then the assumption is not trivial. As I will attempt to illustrate, however, rejecting the assumption that narrow psychological state determines grasped cognitive value is tantamount to abandoning the Fregean conception of reference and content.

The Fregean conception of reference and content maintains that complexes of cognitive values, or *thoughts*, serve as the relata of propositional attitudes; e.g. for Odile and Oscar to have different beliefs is for them to stand in the believes relation to distinct cognitive value complexes. Therefore, if a defender of the Fregean conception denies that narrow psychological state determines grasped cognitive value, then he must hold that it is possible for two subjects to have different beliefs, and yet be in the same narrow psychological state. But in holding this the defender of the Fregean conception undermines his own arguments in favor of positing cognitive values over and above semantic values (i.e. referents and designations). As will be explained in greater detail in section 1.3, cognitive values are posited in the first place to account for our judgments and

behavior with regard to occurrences of sentences; more specifically, cognitive values are posited to explain the *cognitive significance of occurrences*, and thereby account for the phenomenon of opacity.[15] If the defender of the Fregean conception rejects the assumption that narrow psychological state determines grasped cognitive value, he undermines the Fregean conception's account of the phenomenon of opacity.

For example, suppose an understanding subject, say Odile, assents to an occurrence of

(1) Twain is a great author.

yet dissents from an occurrence of

(2) Clemens is a great author.

(Hereafter I will use '(O1)' to refer to the particular occurrence of (1) to which Odile assents, and '(O2)' to refer to the particular occurrence of (2) from which Odile dissents.) The Fregean conception of reference and content accounts for Odile's judgments and behavior by holding that 'Twain' and 'Clemens', as they appear in these occurrences, present distinct cognitive values (for Odile); Odile judges and behaves as she does because these occurrences of 'Twain' and 'Clemens' present distinct cognitive values. Cognitive values are supposed to be in some sense "smack up against Odile's mind" and it is this mental difference which allegedly accounts for the difference in cognitive significance between (O1) and (O2). (Others might call the *cognitive significance* of an occurrence the intuitive *meaning* of the occurrence. I prefer to characterize the cognitive significance of an occurrence, relative to an understanding subject, as the causal effects the perception of the occurrence has on the subsequent mental states and behavior of the subject.) But if cognitive values are to account for the cognitive significance of occurrences and thus

feature in such explanations of judgment and behavior, then--assuming that the causal mechanisms responsible for our judgments and behavior must be physical mechanisms--cognitive values must be represented psychologically; two subjects grasping distinct cognitive values cannot be in the same narrow psychological state. In denying that narrow psychological state determines grasped cognitive value, the defender of the Fregean conception would be granting that the contents of a subject's beliefs, desires, etc. are not relevant for explaining his judgments and behavior, and thus the defender of the Fregean conception would then have to claim that it is the mental states *required* for holding propositional attitudes--not the contents of those attitudes--which are directly relevant to the explanation of judgments and behavior. But if the defender of the Fregean conception is willing to grant this, if he is willing to grant that the content presented by an occurrence does not itself determine the cognitive significance of the occurrence, then why posit cognitive values at all? Why not claim, as do Russellian theorists, that the content of a subject's propositional attitude is not itself directly relevant to the explanation of a subject's judgment and behavior, though the actual brain state of a subject holding an attitude is directly relevant? Why vitiate our ontology with entities that do no explanatory work? If the defender of the Fregean conception denies that narrow psychological state determines grasped cognitive value, then he undermines his own arguments in support of the Fregean conception. Thus Putnam is correct in imputing to the Fregean conception of reference and content the view that narrow psychological state determines grasped cognitive value, and his arguments against this conception are sound.[16]

1.2 The Principle of Semantic Innocence.

The notion of *semantic innocence* derives from this famous passage from Davidson:

> What is strange is . . . that ordinary words for people planets, people,
> tables, and hippopotami in indirect discourse may give up these pedestrian
> references for the exotica. If we could recover our pre-Fregean semantic
> innocence, I think it would seem to us plainly incredible that the words
> 'The Earth moves', uttered after the words 'Galileo said that', mean
> anything different, or refer to anything else, than is their wont when they
> come in other environments. . . . Language is the instrument it is because
> the same expression, with semantic features (meaning) unchanged, can
> serve countless purposes.[17]

In this passage Davidson is criticizing the Fregean analysis of propositional attitude ascriptions, according to which the terms occurring in the that-clause of an attitude ascription do not refer to their ordinary referents, but rather refer to their cognitive values. Let us call the relation that obtains between expressions (tokens and types) and their semantic values (i.e.their referents and designations) the *presents* relation; according to the Fregean conception of reference and content a referring term *expresses* its sense or cognitive value, and *presents* its referent, or semantic value, and similarly a simple declarative sentence *expresses* a thought, and *presents* a proposition.[18] (Note that the Russellian approach has no need for both of these relations, as Russellian theories identify the content expressed by an occurrence with the proposition presented by the occurrence.) Thus according to the Fregean analysis of attitude ascriptions a sentence occurring in a that-clause of an attitude ascription does not present its ordinary semantic value, but instead has the thought it ordinarily expresses as its semantic value; the Fregean analysis maintains that an occurrence of a sentence inside the that-clause of an attitude ascription *presents* what it ordinarily *expresses*. The Principle of Semantic Innocence, however, denies that such shifts in semantic value occur:

<u>The Principle of Semantic Innocence:</u> A declarative sentence occurring inside a that-clause of an occurrence of a propositional attitude ascription presents the same proposition it would have presented if it had occurred on its own, outside of such a that-clause.[19]

Consequently any theory which endorses the Principle of Semantic Innocence is incompatible with the Fregean analysis of attitude ascriptions.

In the passage cited above, Davidson intimates two arguments against the Fregean analysis of attitude ascriptions, and in support of the Principle of Semantic Innocence. First, Davidson suggests that the Fregean analysis is simply counterintuitive: according to the Fregean analysis, referring terms occurring inside the that-clauses of attitude ascriptions refer to their cognitive values rather than their ordinary referents, but intuitively this is not so. It seems that even if it occurs in the that-clause of an attitude ascription, 'Twain' refers to Twain, and not to some "way of thinking of Twain." Examples such as the following make this point abundantly clear:

(a) John: Odile believes that Twain is a great author.
  Mary: Well, he has a sharp wit anyway.

(b) Odile believes that Twain is a great author, but he isn't.

(c) As Odile believes, Twain is a great author.

(d) 1. Odile believes that Twain is a great author.
  2. Twain exists.
  Therefore,
  3. There is somebody who Odile believes to be a great author.

Concerning example (a), it is clear that the pronoun 'he', as it occurs in Mary's utterance is intended to be coreferential with the occurrence of 'Twain' in John's utterance. According to the Fregean analysis, however, 'Twain' as it occurs in John's utterance refers to a cognitive value--a "way of thinking of Twain," and thus 'he' as it occurs in Mary's

utterance must also refer to a cognitive value. But under this interpretation Mary's utterance is nonsense: Mary is not asserting that some (male?) "way of thinking" has a sharp wit.[20] A similar argument applies to example (b). Concerning example (c), someone who utters an occurrence of (c) asserts both that Twain is a great author, and that Odile believes this. In order to assert that Twain is a great author, one must refer to Twain. Thus the occurrence of 'Twain' in (c) must refer to Twain. On the other hand, if one is to ascribe a belief about Twain to Odile, then, according to the Fregean analysis, one must specify some "way of thinking of Twain." Consequently, 'Twain' as it appears in an occurrence of (c) must refer to a cognitive value. But surely this one occurrence of 'Twain' does not refer to both Twain and a cognitive value.[21] And finally, the "exportation" argument presented in example (d) is intuitively valid, but how can its validity be accounted for if 'Twain' as it appears in premise 1 refers to a "way of thinking of Twain" and not to Twain? If premise 1 does not present a proposition concerning the man Twain, how can premise 3, which does present a proposition concerning the man Twain, follow from premises 1 and 2? Examples such as these weigh heavily against the Fregean analysis of attitude ascriptions and in favor of the Principle of Semantic Innocence.

The second argument intimated in the above citation from Davidson concerns *the productivity of language and thought*. Despite our limited cognitive capacities, human beings have the ability to think any one of a huge number of thoughts, and, correspondingly, the ability to understand, or determine the proposition presented by, any one of an enormous number of sentences. For example, it is unlikely that the reader has encountered the sentence, 'Clinton has a pet armadillo' before now, but even so you can understand this sentence; you know what proposition an occurrence of this sentence would present, and you know what the truth conditions of an occurrence of this sentence would be. It is widely thought, and correctly so on my view, that this ability of normal speakers can be explained only if language is compositional; you can determine the proposition

presented by an occurrence of 'Clinton has a pet armadillo' because you (tacitly) know the semantic values of each of the terms appearing in the sentence, and from your perception of the token of this sentence you can determine the syntactic structure of the sentence, and finally you know (tacitly) compositional rules for determining the proposition expressed from the basic semantic values and the syntactic structure. If this kind of account of the productivity of language and thought is correct, there must be substantial limits on the amount of basic knowledge that speakers have. That is, this explanation is plausible only in so far as it attributes to ordinary speakers knowledge of a finite number of compositional rules and knowledge of the semantic values of a finite number of terms. The difficulty with the Fregean analysis of attitude ascriptions is that it requires that ordinary speakers have an immense store of such basic knowledge.

Davidson has argued, and conclusively so in my opinion, that if the Fregean analysis of attitude ascriptions is understood as asserting that each referring term is infinitely ambiguous, then the Fregean analysis will be unable to account for the *learnability* of natural languages.[22] Consider for example 'Portland'. This one inscription type corresponds with two distinct words: there is the word 'Portland*' which refers to Portland Maine, and there is the word 'Portland**' which refers to Portland Oregon. The fact that 'Portland*' and 'Portland**' are homonyms with the same spelling is something of an accident. If 'Twain' is similar to 'Portland' in that there are really distinct words, 'Twain*' which refers to Twain, and 'Twain**' which refers to a cognitive value, then English would contain infinitely many primitive words and therefore would be unlearnable. The problem arises from iterated attitude ascriptions. It is an empirical fact that if a competent speaker can understand an occurrence of

(3) Odile believes that Twain is a great author.

then he can also understand an occurrence of

(3') Odile believes that she believes that Twain is a great author.

To what would 'Twain' in an occurrence of (3') refer? Since an occurrence of (3') states that Odile holds the attitude of belief toward a thought which is itself about a thought which is about Twain, it seems that 'Twain' in an occurrence of (3') must refer to a "way of thinking of a way of thinking of Twain," or what Frege called a "secondary sense."[23] (This is also the analysis offered by Church, whom Davidson is criticizing.[24]) The problem is of course that attitude ascriptions can be embedded indefinitely, and any competent speaker who understands (3') can also understand further embeddings. Hence if 'Twain' is ambiguous in the way that 'Portland' is ambiguous, ordinary speakers will be required to know that 'Twain*' refers to Twain, that 'Twain**' refers to a "way of thinking" of Twain, that 'Twain***' refers to a "way of thinking of a way of thinking of a way of thinking" of Twain, etc. The same considerations apply, of course, to all terms. Consequently, if the Fregean analysis is understood as attributing genuine ambiguity to all terms, then the Fregean analysis of attitude ascriptions is committed to ordinary speakers having far too much basic knowledge.

As Davidson points out, however, the Fregean analysis of attitude ascriptions need not be understood in this way. The Fregean analysis can be understood as claiming not that all referring terms are ambiguous in the way that 'Portland' is ambiguous, but rather as claiming that all referring terms are *systematically ambiguous*. That is, the Fregean theorist can avoid the result that speakers must have basic knowledge of the semantic values of an infinite (or at least very large number) of primitive terms by claiming that all terms are in a way indexical. That is, the Fregean analysis can be understood as maintaining that the semantic value of an occurrence of a referring term depends in part upon where the term appears in the sentence. Understood in this way 'Twain', for example, is not "infinitely ambiguous" anymore than 'you' is infinitely ambiguous.

Normal speakers manage to learn how to speak and understand English even though 'you' can be used to refer to an infinite number of things. So perhaps 'Twain', though not *ambiguous*, is *protean*, and thus occurrences of it can have any one of a large number of semantic values. Perhaps what ordinary speakers who understand occurrences of (3), (3') etc. know is not the semantic values of a huge number of primitive *terms*, but rather one general *rule* for determining the semantic value of 'Twain' relative to different sentential environments.

Though this interpretation of the Fregean analysis is not objectionable on the grounds that it is committed to their being too many *primitive terms*, it is nonetheless objectionable on the grounds that it is committed to an ordinary speaker's having (tacit) knowledge of too many *primitive semantic values*.[25] According to the Fregean analysis of attitude ascriptions the semantic value of a term, relative to a context of utterance, varies depending upon whether the term occurs inside a that-clause, or outside a that-clause: outside of a that-clause a term refers to its ordinary referent, yet inside a that-clause a term refers to its cognitive value, and inside an embedded that-clause a term refers to a secondary cognitive value which determines a primary cognitive value, etc. Recall, however, that the cognitive value presented by a referring term is something like a "way of thinking" of the term's referent. Consequently, since there are, I assume, many different "ways of thinking about Twain," there is no unique cognitive value presented by an occurrence of 'Twain'. Rather, the cognitive value presented by even a single occurrence of 'Twain', *varies intersubjectively*: the same occurrence of a referring term will present distinct cognitive values for different subjects.[26] (I argue in support of this assumption in section 2.2.) As a result, the Fregean analysis is committed to the claim that the semantic values of referring terms which appear inside of that-clauses can only be determined relative to particular subjects (or even particular subjects at particular times.) Thus, under the Fregean analysis, a given occurrence of a referring term may refer to the

27

referent itself, to any one of a huge number of "ways of thinking of its referent," or to "ways of thinking of ways of thinking of the referent," etc.

Of the many potential semantic values of an occurrences of a referring term, how is it determined which one is expressed by a term occurring in the that-clause of an ascription? What, for example, is the semantic value expressed by 'Twain' as it appears in an occurrence of (3)? Since cognitive values serve as the contents of propositional attitudes, and (3) states that Odile holds the attitude of belief toward a particular thought, an occurrence of (3) is true only if 'Twain', as it appears in (3), refers to a "way of thinking of Twain" which is a constituent of a content toward which Odile holds the attitude of belief. But this is problematic: any competent speaker of English who has mastered the meanings of 'Twain' and 'Odile'--whatever this amounts to--can understand an occurrence of (3), and therefore know what it would be for an occurrence of (3) to be true. But do all such competent speakers know anything at all about the "way in which Odile thinks of Twain"? Is it at all plausible to suppose that these speakers know (tacitly) that the semantic value of 'Twain' as it appears in an occurrence of (3) is W where W is a particular "way in which Odile thinks of Twain"?

Once a speaker masters the meaning of 'Twain' he is able to produce myriad thoughts and sentences about Twain; he would be able to sit in the philosopher's armchair and produce, and understand, myriad sentences containing 'Twain'. Among the myriad sentences our subject would be capable of producing and understanding would be many attitude ascriptions. For example, if our subject, a normal speaker of English, has mastered the meaning of the relevant terms, he would be able to produce, and understand, occurrences of the following sentences:

(5) Nixon believed that Twain was a quarterback.

(6) Bush believed that Aristotle was identical to Twain.

(7) Nixon believed that Dostoevsky believed that Twain played quarterback.

Our subject can understand, or know what proposition is presented by, occurrences of these sentences only if he knows (tacitly) what semantic values are presented by 'Twain' as it appears in these occurrences, and the Fregean analysis of attitude ascriptions maintains that the semantic values presented by these appearances of 'Twain' are various "ways of thinking of Twain": 'Twain' as it appears in an occurrence of (5) has W' as its semantic value, where W' is some particular "way in which Nixon thought of Twain"; 'Twain' as it appears in an occurrence of (6) has W" as its semantic value, where W" is a particular "way in which Bush thinks of Twain"; and finally 'Twain' as it appears in an occurrence of (7) must have W"' as its semantic value, where W"' is a particular "way in which Nixon thinks of a particular way in which Dostoevsky thinks of Twain." All of these analyses, however, are implausible. Surely in mastering the meaning of 'Twain' and thereby acquiring the ability to formulate and understand myriad sentences containing 'Twain', our subject does not acquire the ability to refer to such exotic entities as W', W" and W"'. Surely a speaker can understand an occurrence of (6) even if he has no idea at all as to "the way in which Bush thinks of Aristotle." What if, as would seem to be likely, Bush thinks of Aristotle in many different "ways"? Which one of them is referred to by a given occurrence of (6), and, more importantly, how does a speaker who knows nothing about Bush's Aristotle thoughts determine that this, rather than that, "way of thinking" is the referent of 'Aristotle'? Moreover, since our subject can, in principle, iterate 'believes that' *ad infinitum,* it seems that the Fregean analysis of iterated attitude ascriptions can account for our ability to understand occurrences of such ascriptions only if we have (tacit) knowledge of, and therefore the ability to refer to, an indefinitely large number of these exotic entities. But to attribute to a normal speaker who has only just mastered the meaning of 'Twain' such knowledge and ability is to attribute to him (tacit) knowledge of far too many primitive semantic values. The problem here is not that there are too many primitive *terms,* but rather that there are too many *referents.*

Another problem with the Fregean analysis of attitude ascriptions, which is related to the problem discussed above, is that in many cases the semantic values required by the Fregean analysis could not exist. For example, a normal speaker of English who has mastery of the relevant terms is able to formulate the following sentences, and know what propositions occurrences of them would present:

(8n) Aristotle did not believe that Twain was a great author.

(9) Nobody believes that Twain played quarterback.

According to the Fregean analysis, however, in order for our subject to know the truth conditions of an occurrence of (5n), our subject would have to know that 'Twain' as it appears in the occurrence refers to "Aristotle's way of thinking of Twain." This is problematic because, of course, Aristotle did not think of Twain in any "way" at all and thus the Fregean analysis seems to incorrectly predict that an occurrence of (8n) would be meaningless, or lack a truth value, since it would suffer from reference failure. Concerning occurrences of (9), the Fregean analysis of attitude ascriptions maintains that in order to know the truth conditions of an occurrence of (9), our subject would have to know that 'Twain' in (9) referred to W"", where W"" is "the way nobody thinks of Twain," but this is nonsense. (Also, there is no straightforward way in which (9) can be analyzed as involving "hidden" quantification over cognitive values.[27])

These consequences run counter to known facts. Once one knows that 'Twain' refers to Twain (whatever possessing this knowledge might amount to), one can formulate myriad sentences such as (5)-(9), and, given an appropriate store of abilities and basic knowledge, one can know the propositions presented by, and therefore the truth conditions of, occurrences of these sentences. To maintain that in order know the truth conditions of occurrences of (5)-(9) a subject must know that 'Twain' refers to this or that "way of

thinking of Twain", or this or that "way of thinking of this or that way of thinking of Twain," is to attribute to ordinary speakers basic knowledge of far too many primitive semantic values. Consequently the Fregean analysis of attitude ascriptions cannot account for, and is incompatible with, the productivity of language and thought.

1.3 The Achilles' Heal of the Russellian Approach: The Arguments and Problems Posed by the Phenomenon of Opacity.

The resurgence of Russellian theories is due to the arguments presented above (or very similar arguments) in support of the Principle of Direct Reference and the Principle of Semantic Innocence. Though none of these arguments is conclusive, I find these arguments compelling, and thus I too am motivated to formulate and defend a Russellian theory. The central difficulty facing Russellian theories is accounting for the phenomenon of opacity. As was remarked above, Frege's theory of sense and reference was developed specifically to account for this phenomenon, and thus it is not surprising that Russellian theories, which by definition deny several fundamental tenets of Frege's theory, encounter some difficulty in accounting for this phenomenon.

The arguments and problems posed by the phenomenon of opacity are perhaps best introduced as objections to what I will call *The Naive Russellian Theory*. In addition to the principles of direct reference and semantic innocence, the Naive Russellian Theory maintains the following two principles:

> The Principle of Full Articulation: The proposition presented by an occurrence of a declarative sentence is wholly determined by (a) the semantic values--the referents and designations--of the phonetically or orthographically realized terms occurring in the sentence, (b) the syntactic structure of the sentence.[28]

> The Principle of Propositional Truth: The truth conditions of an
> *occurrence* of a declarative sentence are identical to, and determined by,
> the truth conditions of the *proposition* presented by the occurrence; i.e. no
> two occurrences which present the same proposition can have distinct truth
> conditions. (Note that as I use 'proposition', it is not definitive of a
> proposition p that p determine the truth value of an occurrence which
> presents p. That is, the Principle of Propositional Truth is not entailed by
> my account of propositions. See note 2.)

And finally, the Naive Russellian Theory endorses a binary analysis of propositional

attitudes and propositional attitude verbs: propositional attitudes--the actual mental

phenomena--are binary relations between agents and Russellian propositions and,

furthermore, occurrences of propositional attitude verbs designate these binary

relations.[29]

Before explicating how the arguments and problems posed by the phenomenon can

be invoked against the Naive Russellian Theory, a few remarks of clarification concerning

the Principle of Propositional Truth are in order. The above statement of the Principle of

Propositional Truth presupposes that *occurrences*--tokens in contexts--and *propositions*

both have truth conditions. The presumption that propositions have truth conditions is, I

assume, unproblematic. (At least it is familiar.) But what does it mean to say of an

occurrence that it is true, or false? Suppose that an ordinary English speaker, Odile,

perceives an occurrence of

(1) Twain is a great author.

Under certain conditions Odile will, as a result of her perception of the occurrence, be led

to make a judgment concerning truth value. That is, upon hearing, seeing, or feeling the

occurrence she might utter "That's true," or "That's false," or even "That's neither true nor

false." Whatever judgment concerning truth value Odile might make, I assume that her

judgment is either correct or incorrect; i.e. I assume that if Odile utters an occurrence of

'That's true', then that utterance has a determinate truth value. What determines whether or not Odile's utterance of 'that's true' is true? The traditional answer to this question, which is encapsulated in the Principle of Propositional Truth, is that Odile's utterance of "That's true" is true just in case the proposition presented by the occurrence of (1) is true. Thus, according to the traditional view, the demonstrative 'that' appearing in Odile's utterance of 'That's true' refers to the proposition presented by the occurrence of (1).

What I am attempting to draw the reader's attention to is that the traditional view encapsulated in the Principle of Propositional Truth is a theory concerning the truth conditions of the judgments ordinary speaker's make in response to their perception of occurrences of declarative sentences. It is not a *datum* that the correctness of a judgment of truth value in response to an occurrence is determined by (i) the syntactic structure of the sentence, and (ii) the semantic values--the referents and designations--of the terms appearing in the occurrence. That is, it is not a datum that an intuitive judgment of truth value made in response to an occurrence is about the proposition the occurrence presents *only*. It is well known that an occurrence typically *conveys*--in part via pragmatic mechanisms--much more information than it *semantically encodes* or *presents*. There is no pre theoretic reason to follow the traditional view in assuming that this additional information which is conveyed via pragmatic mechanisms, rather than semantically encoded, plays no part in determining the correctness of the judgments of truth value speakers make in response to occurrences. In taking pains to point this out, I am of course tipping my hand: In the end I will reject the Principle of Propositional Truth. That said, let us return to the problems and arguments posed by the phenomenon of opacity.

Though it is often alleged that the Naive Russellian Theory is refuted by arguments which are based upon the phenomenon of opacity, it is rarely noticed that there are two kinds of arguments from opacity: there are *epistemological* arguments from opacity and there are *semantic* arguments from opacity. *Epistemological* arguments from opacity are concerned with the epistemological component of the Naive Russellian

Theory; they are directed against the Naive Theory's claim that Russellian propositions are the contents of propositional attitudes. The *semantic* arguments from opacity, on the other hand, are concerned with the semantic component of the Naive Russellian theory; they are designed to show that the Naive Russellian Theory makes incorrect predictions concerning the truth conditions of occurrences of propositional attitude ascriptions. Drawing this distinction between the two kinds of arguments is crucial to my defense of the Russellian approach, as I maintain that the epistemological arguments from opacity against the Naive Russellian theory are unsound, while the semantic arguments from opacity force the Russellian theorist to abandon the Naive Russellian Theory. (More specifically, I maintain that the semantic arguments from opacity force Russellian theorists to reject the Principle of Propositional Truth.)

I begin by explicating an example of an epistemological argument from opacity. Suppose that a sincere, normal, understanding subject, say Odile, assents to an occurrence of

(1) Twain is a great author.

yet dissents from an occurrence of

(2) Clemens is a great author.

(I will continue to call these occurrences '(O1)' and '(O2)', respectively.) Because the Naive Russellian Theory endorses the Principle of Direct Reference, it must maintain that the relevant tokens of 'Twain' and 'Clemens' present the same content, viz. Twain, the man. Consequently, since (O1) and (O2) contain (tokens of) the same predicate and have (or are instances of) the same syntactic structure, The Naive Russellian Theory dictates that (O1) and (O2) express the same content, viz. the Russellian proposition containing

Twain, the man, and the property of being a great author as constituents. Let us call this Russellian proposition, 'P'. The epistemological arguments assume that it follows from Odile's assent to (O1) that she holds the attitude of belief toward the content expressed by this occurrence, and that it follows from Odile's dissent from (O2) that she does not hold the attitude of belief toward the content expressed by this occurrence. More generally, the epistemological arguments from opacity assume the following two principles:

> The Epistemological Principle of Assent: If, at time t, a normal, sincere, understanding subject assents to an occurrence of a declarative sentence Σ, then he holds the attitude of belief toward the content expressed by that occurrence of Σ, at time t.

> The Epistemological Principle of Dissent: If, at time t, a normal, sincere, understanding subject dissents from an occurrence of a declarative sentence Σ, then he does not hold the attitude of belief toward the content expressed by that occurrence of Σ, at time t. (Note that dissenting from an occurrence of a sentence is to be distinguished from assenting to the negation of an occurrence of a sentence.)

It follows from the Principle of Assent that at the time of her assent Odile holds the attitude of belief toward the content expressed by (O1); or more formally, it follows from Odile's assent to (O1) at t that

(A) Believes <Odile, P> at t.[30]

Furthermore, it follows from the Principle of Dissent that at the time of her dissent Odile does not hold the attitude of belief toward the content expressed by (O2); or more formally, it follows from Odile's dissent from (O2) at t' that

(B) ¬(Believes <Odile, P> at t').

35

Assuming that Odile does not "change her mind" concerning (O1) during the interval between t and t', it also follows that

(A') Believes <Odile, P> at t.

But (A') and (B) are contradictories; Odile cannot both hold the attitude of belief toward P at t', and not hold the attitude of belief toward P at t'. Thus the Naive Russellian Theory, coupled with some plausible assumptions concerning what follows from Odile's assenting and dissenting, leads us to a contradiction. Therefore, the argument concludes, at least one of the tenets of the Naive Russellian Theory must be rejected.

Another problem for the Naive Russellian Theory, which is closely associated with the epistemological arguments from opacity, concerns *the cognitive significance of occurrences*. As was touched upon in section 1.1., the Fregean conception of reference and content maintains that Odile assents to (O1), yet dissents from (O2) *because* these occurrences express distinct contents, and they express distinct contents because 'Twain' as it appears in (O1), and 'Clemens' as it appears in (O2), express distinct cognitive values. Thus, according to the Fregean conception of reference and content, Odile's behavior with regard to (O1) and (O2) is caused by, or can at least be explained by appeal to, the contents expressed by (O1) and (O2). The Naive Russellian Theory, however, cannot avail itself to this sort of account of the cognitive significance of occurrences because it maintains that two occurrences which have distinct cognitive significance may present the same content--(O1) and (O2) are examples of such occurrences. But surely, the objection continues, it is the *content* expressed by an occurrence which is responsible for the *cognitive significance* of the occurrence? Surely Odile assents to (O1) *because* she holds the attitude of belief toward the content expressed by this occurrence, and surely she dissents from (O2) *because* she does not hold the attitude of belief toward the content

36

expressed by this occurrence. Therefore, it seems that The Naive Russellian Theory cannot account for the cognitive significance of occurrences.

I now turn to explicating the semantic arguments from opacity against The Naive Russellian Theory. The central difference between the epistemological arguments and the semantic arguments concerns what follows from the subject's assenting to and dissenting from the relevant occurrences. Where the epistemological arguments presuppose the above epistemological principles of assent and dissent, the semantic arguments from opacity presuppose the following semantic principles of assent and dissent:

> The Semantic Principle of Assent: If a normal, sincere, understanding subject assents to an occurrence of a declarative sentence Σ in a context c, then an occurrence of $\ulcorner$N believes that $\Sigma\urcorner$, where N refers to the subject, is true in any context c', where c' is *semantically similar to c with regard to* Σ *and* $\ulcorner$N *believes that* $\Sigma\urcorner$. (A context c is *semantically similar to a context c' with regard to sentences* Σ *and* Σ' iff for every term τ appearing in Σ or Σ', τ presents the same semantic value relative to both c and c'.[31])

> The Semantic Principle of Dissent: If a normal, sincere, understanding subject dissents from an occurrence of a declarative sentence Σ in a context c, then an occurrence of $\ulcorner$N does not believe that $\Sigma\urcorner$, where N refers to the subject, is true in any context c' where c' is *semantically similar to c with regard to* Σ *and* $\ulcorner$N *does not believe that* $\Sigma\urcorner$. (Again note that dissenting from an occurrence of a sentence is to be sharply distinguished from assenting to the negation of an occurrence of a sentence.)

Assuming these principles, it follows from Odile's assent to (O1) in c that an occurrence of

> (3) Odile believes that Twain is a great author.

is true in some semantically similar context c'. Furthermore, it follows from Odile's dissent from (O2) in c that an occurrence of

(4n) Odile does not believe that Clemens is a great author.

is also true in c'. (For the sake of clarity I am making a few simplifying assumptions here; a similar argument could be formulated without the benefit of these assumptions.) The Naive Russellian Theory dictates, however, that for any context in which an occurrence of (3) is true, an occurrence of

(4) Odile believes that Clemens is a great author.

is also true. (If (3) expresses proposition P in c, then, by the principles of direct reference, semantic innocence, and full articulation, (4) must also express P in c, and thus, by the Principle of Propositional Truth, occurrences of (3) and (4) in c must have the same truth conditions.[32]) Consequently, since (3) is true in c', it follows that (4) is true in c'. But (4) is simply the negation of (4n), and it was previously determined that (4n) is true in c'. But a sentence and its negation cannot both be true relative to the same context, and consequently, the argument concludes, the Naive Russellian Theory must be rejected.

Yet another problem for The Naive Russellian Theory, which is related to the semantic arguments from opacity and the problem of cognitive significance, concerns preserving our untutored intuitions concerning the truth conditions of common belief\desire explanations of behavior. "Folk psychology" dictates that a subject's judgments and behavior is determined by, or at least can be explained by, what he believes, desires, fears, etc. For example, our untutored intuitions dictate that, in many contexts c, an occurrence of

(10a) Odile desires that Twain come to the party

and

(10b) she believes that if wine is served, then Twain will come to the party,

and that's why

(10c) Odile is going to serve wine.

is true in c, while an occurrence of

(10a) Odile desires that Twain come to the party.

and

(10b') she believes that if wine is served, then Clemens will come to the Party,

and that's why

(10c) Odile is going to serve wine.

is false in c. The difficulty presented by such folk explanations is not merely that of accounting for why the occurrences of the sentential constituents (10b) and (10b') can have distinct truth conditions--that is just the problem posed by the semantic arguments from opacity. Rather the problem presented by such folk explanations is to account for why what is conveyed by occurrences of (10a) and (10b) might serve as an explanation of what is conveyed by (10c), while what is conveyed by occurrences of (10a) and (10b') might not serve as an explanation of what is conveyed by (10c). The problem is not accounting for the truth conditions of the sentential constituents of occurrences of (10a-b-c) and (10a-b'-c), rather the problem is accounting for the truth conditions of occurrences of (10a-b-c) and (10a-b'-c) *as whole sentences.*

Theories which utilize the Fregean conception of referenced and content and the Fregean analysis of attitude ascriptions can, it seems, preserve our untutored intuitions concerning the truth conditions of such folk explanations of behavior. Such theories

maintain that content determines cognitive significance, i.e. to the extent that the judgments and behavior of a subject can be explained by the propositional attitudes of the subject, it is the contents of the subject's propositional attitudes which are explanatorily relevant. Thus theories which utilize the Fregean conception of reference and content and the Fregean analysis of attitude ascriptions can account for the explanatory relation which holds between what is conveyed by occurrences of (10a-b) and what is conveyed by occurrences of (10c), yet does not hold between what is conveyed by occurrences of (10a-b') and what is conveyed by occurrences of (10c). More precisely, Fregean theories can account for truth of (10a-b-c) on the grounds that this occurrence presents something of the following form:

(a) X desires content C
(b) X believes (CV(E occurs) → C is realized)
Therefore,
(c) X acts so as to make E occur.

('X' in the above formula is a schematic letter open for replacement by names for subjects; 'C' is a schematic letter open for replacement by terms referring to Fregean thoughts; 'E' is a schematic letter open for replacement by expressions designating potential events. 'CV( )' denotes a function from semantic values to cognitive values; i.e. from propositions to the thoughts which determine them. I here overlook the fact that there could be no such function--for every proposition there are many thoughts which determine it.) An occurrence of (10a-b-c) presents something of this form because, according to the Fregean analysis, the that-clause of an occurrence of (10a) presents the same content as the consequent in the that-clause of an occurrence of (10b). An occurrence of (10a-b'-c), however, does not present something which is of the above form: according to the Fregean analysis of attitude ascriptions, the that-clause of an occurrence of (10a) and the consequent in the that-clause of an occurrence of (10b') present distinct contents. Thus

theories which utilize the Fregean conception of reference and content and the Fregean analysis of attitude ascriptions can account for the truth of occurrences of (10a-b-c), and falsity of occurrences of (10a-b'-c): Occurrences of (10a-b-c) present something of the appropriate form, while, because occurrences of (10b) and (10b') present distinct propositions, occurrences of (10a-b'-c) do not present something of the appropriate form.

The Naive Russellian Theory, however, is unable to distinguish the truth conditions of common belief\desire explanations such as (10a-b-c) and (10a-b'-c). The Naive Russellian Theory endorses the Principle of Semantic Innocence, and thus denies that the that-clauses of occurrences of (10b) and (10b') present distinct contents; rather, the Naive Russellian Theory maintains that the that-clauses of occurrences of (10b) and (10b') have the same Russellian proposition as semantic value. Therefore, since it endorses the Principle of Propositional Truth, The Naive Russellian Theory cannot distinguish the truth conditions of occurrences of (10a-b-c) and (10a-b'-c). Moreover, according to the Naive Russellian Theory occurrences of the that-clause of (10a) and the consequent of occurrences of the that-clause of (10b') express (and present) the same content, and thus occurrences of (10a-b-c) and (10a-b'-c) both express (and present) something of the appropriate form. Consequently, the Naive Russellian theorist cannot account for our untutored intuitions concerning the truth values of occurrences of (10a-b-c) and (10a-b'-c) by claiming that occurrences of (10a-b-c) present something of the appropriate form, while occurrences of (10a-b'-c) do not present something of this form.

Introduction and Chapter 1 Notes

[1] Excerpt from Russell's letter to Frege, December 12th, 1904. Excerpts of Frege and Russell's correspondence are printed in Salmon, Soames (eds.) 1988.

[2] As I use 'proposition', the identity of a proposition presented by an occurrence of a declarative sentence is wholly determined by (i) the logical form of the sentence, and (ii) the semantic values presented by the occurrence, i.e. the entities referred to or designated by the terms appearing in the occurrence. Russellian propositions are thus structured abstract objects which "bear" the properties of truth and falsity. More specifically, propositions can be identified with interpreted phrase structure markers (syntactic trees) at the level of logical form (LF). (Both syntactic trees and interpreted syntactic trees are abstract objects.) An interpreted tree is obtained from a syntactic tree by "replacing" the lexical items appearing in the lowermost nodes of the syntactic tree with the semantic values of those lexical items. (The semantic values of context sensitive terms can only be assigned relative to an assignment function, which I assume is supplied by the context of utterance.) For example, the syntactic tree at the level of LF for the sentence 'Twain smokes' is represented by the following diagram:

Syntactic Tree 1.
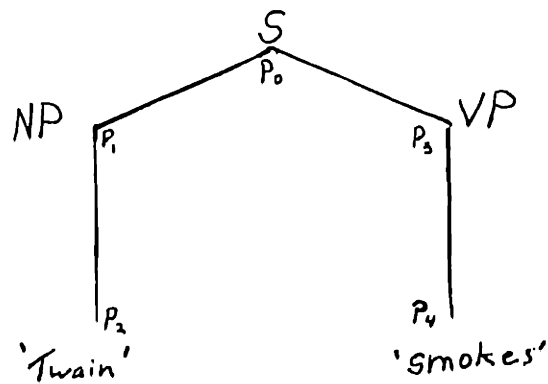
An interpreted tree is now obtained by "replacing" 'Twain' on p2 with Twain and 'smokes' on p4 with the property of being a smoker. This interpreted syntactic tree 1 is picturesquely depicted in the following diagram:

Interpreted Tree 1.
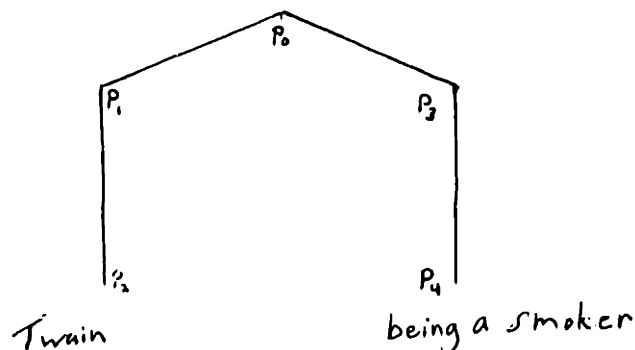
Note that Interpreted Tree 1 would also be derived from the Syntactic Tree corresponding to 'Clemens smokes'; this is because both 'Twain' and 'Clemens' will be "replaced' by Twain, the man. Moreover the syntactic trees corresponding to 'Odile believes that Twain smokes' and 'Odile believes that Clemens smokes' also determine the same interpreted syntactic tree. (What exactly is it to "replace" 'Twain', an abstract object which serves as a kind of constituent of the syntactic tree, with *Twain*, a now dead man, in a syntactic tree? Syntactic trees can be defined in set theoretic terms; every "fork" in a syntactic tree at the level of logical form is identified with an ordered pair, where the "left prong of the fork" is identified with the "left member" of the ordered pair, etc. Thus the syntactic tree for the sentence 'every man smokes' can be identified with the set, <'every' <'man', 'smokes'>>, and the syntactic tree represented by the above diagram can be identified with the ordered set <'Twain', 'Smokes'>. To "replace" the lexical items in the lowermost nodes of a syntactic tree with the semantic values of those lexical items is simply to obtain from an ordered set which has lexical items as its basic elements another set with the same structure which has as basic elements the semantic values of the original basic elements. But what is it for *Twain*, a man long dead, to be an element of such a set? Indeed, what exactly is a *property*? I will not delve into such metaphysical questions here. Suffice it to say that whatever it is that we, in 1993, refer to when using 'Twain' and designate when using 'smokes', these are the elements of interpreted syntactic tree 1.)

The above illustrates that the Russellian approach to propositions is a species of, and not an alternative to, the so called "Interpreted Logical Form" view of propositions. (See Higginbotham, 1991.) In order to avoid the problems associated with opacity, Higginbotham maintains that an occurrence of a that-clause refers to itself, or its syntactic tree, where the semantic values of the syntactic items appearing in the trees are added to the trees in special cases only. According the Russellian approach, on the other hand, an occurrence of a that-clause always refers to its fully interpreted logical form (or syntactic tree); an occurrence of a that-clause always refers to what one derives via the procedure adumbrated above.

3 I take the *Russellian approach* to be what all *Russellian theories* have in comr..un. A *theory* of propositional attitudes and propositional attitude ascriptions is *Russellian* if and only if it respects the principles of Direct Reference and Semantic Innocence.

4 Kaplan 1989b, pp. 568-9.

5 I follow Kaplan in my usage of 'occurrence': an occurrence of a term is a token in a context of utterance, or a context of inscription. I believe that Kaplan introduces this term in Kaplan 1989a.

6 Dummett has claimed that the causal theory of reference is "only a theory about the senses of . . . names rather than one which replaces sense by something different.," (Dummett, "Frege's Distinction Between Sense and Reference," in Dummett, 1978. Here Dummett seems to be *equating* sense with that which determines reference, and if we follow Dummett in equating these things,then his point correct. It is, however, a mistake to *equate* sense with that which determines reference, for it is essential to Frege's

conception of sense that senses also serve as the contents expressed by occurrences of terms.

7 The reason for the, perhaps excessive, use of "scare quotes" around such phrases as 'way of thinking of the referent' is that I do not know what such "ways of thinking" are. I want to remind the reader that these phrases require further explanation. (I suspect that, in most cases, saying "X thinks of Y in way Z" is another of saying that X *thinks that* Y has some class of properties; this interpretation, however, will not suite the Fregean theorist's purposes--he is trying to analyze *thinking that* in terms of "ways of thinking.")

8 In "Demonstratives," Kaplan advances the relatively conservative thesis that demonstratives and indexicals are directly referential. According to the Russellian approach defended here, however, all occurrences of referring terms--proper names, indexicals, demonstratives, deictic pronouns, and perhaps natural kind terms--are directly referential. (Note that definite descriptions are not directly referential. Following Russell, I will say that an occurrence of a definite description which manages to "pick an object out" *denotes* the object it "picks out.")

9 Though it will not be discussed here, I also endorse what might be called a Principle of Direct Predication which maintains that the content expressed by an occurrence of a predicate is the property designated by the predicate and not whatever it is that determines the property.

It is interesting to note that such a Principle of Direct Predication has not received the attention that the Principle of Direct Reference has received. This is primarily because the notion of sense, "mode of presentation," or a "way of thinking," is superfluous in the case of predicates. For example, 'renates are cordates' does not express an "obvious truth," yet Fregean theorists need not posit senses or modes of presentation in order to explain this fact; it can simply be claimed that 'renate' and 'cordate' designate different, yet coextensive, properties; they express distinct contents. Similarly, senses or modes of presentation need not be posited in order to explain the fact that a subject might assent to an occurrence of 'Turtles are renates', yet dissent from an occurrence of 'Turtles are cordates'; since the two predicates designate distinct properties, the two sentences express distinct propositions. There is no reason to posit senses, or more generally cognitive values, for non referring terms. (A historical point: Frege himself had need of senses for predicates because he nearly identified the concept expressed by a one-place predicate with its extension. A concept for Frege is, we might say, a "non-saturated" extension. Hence, for Frege occurrences of 'renate' and 'cordate' would designate the same concept, and he needs senses in order to explain the fact that 'renates are cordates' does not express an "obvious truth." My point is that, unless one accepts the rather dubious saturated/unsaturated distinction, one has no reason whatsoever for positing senses in addition to concepts and extensions. For a discussion of his views concerning concepts, see Frege, "Comments on Sense and Meaning," in Frege 1979.

10 Kripke 1972, and Putnam 1975.

11 In a famous footnote, Frege states

> In the case of an actual proper name such as 'Aristotle' opinions as to the
> sense may differ. It might, for instance, be taken to be the following: the
> pupil of Plato and teacher of Alexander the Great. Anybody who does this
> will attach another sense to the sentence 'Alexander was born is Stagira'
> than will a man who takes as the sense of the name: the teacher of
> Alexander the Great who was born in Stagira. So long as the [referent]
> remains the same, such variations of sense may be tolerated, although they
> are to be avoided in the theoretical structure of a demonstrative science and
> ought not to occur in a perfect language.

(See, "On Sense and Reference," in Frege 1966, p. 58, footnote.) In this passage Frege seems to have overlooked that senses often serve as referents; any difference in sense will, in some contexts, result in a difference of referent. Thus it would seem that even differences of sense ought not be "tolerated."

12 In the ensuing discussion I do not consider the "cluster theory" and thus a proponent of the cluster theory might object that I have given descriptivist theories of cognitive value short shrift. In response to such an objector I would point out, first, that my purpose here is not so much to refute descriptivist theories, as to illustrate how Kripke's arguments against such theories serve to support the Russellian approach to propositional attitudes and propositional attitude ascriptions. Second, it seems to me that though the cluster theory is at least a plausible theory of reference determination, it is an extremely implausible theory of content. It is surely implausible to suppose that 'Aristotle' as it occurs in 'Aristotle taught Alexander the Great' expresses (with varying weights) the content of every description the speaker associates with 'Aristotle'. Note that all that is important for supporting the Principle of Direct Reference is that the content of the cluster of descriptions the speaker associates with a term cannot both determine the referent of the term and serve as the content presented by occurrences of the term. This is why I do not bother to summarize Kripke's arguments against the cluster theory qua a theory of reference determination.

13 Kripke 1972, p. 30.

14 Putnam explains that no narrow psychological state "presupposes the existence of any individual other than the subject to whom that state is ascribed" (Putnam 1975, p. 220). Putnam points out that the only motivation for restricting talk of psychological states to talk of such narrow states is the assumption that "psychological states in the narrow sense have a significant degree of causal closure (so that restricting ourselves to psychological states in the narrow sense will facilitate the statement of psychological laws)" (Putnam 1975, p. 221). Putnam is rightly skeptical of this claim; it seems that there may be laws of psychology which appeal to wide psychological states. (E.g. "If x is jealous of y, then x will . . . . toward y.") But the rather strong claim that there are no wide psychological

states which can play a role in psychological laws and in psychological explanations should be distinguished from the much more plausible claim that if two subjects behave in different ways, then they must be in different narrow psychological states. If physicalism is true, and I assume that it is, then behavior is caused by some kind of physical mechanism; since psychological states are individuated by the kind of behavior the instances of the states cause, different kinds of behavior must be accompanied by different narrow psychological states. To reject this claim is to reject physicalism.

15 First, note that I speak of the *phenomenon of opacity*, rather than of "opaque contexts," "opaque positions," or "opaque constructions." The phenomenon of opacity is, roughly, the phenomenon of a normal understanding subject's assenting to an occurrence of $\Sigma(\alpha)$, yet dissenting from an occurrence of $\Sigma(\beta)$, where $\Sigma(\ )$ is a sentential function and α and β are coreferring terms. (Thus $\Sigma(\alpha)$ and $\Sigma(\beta)$ are sentences which differ only in that $\Sigma(\alpha)$ has α in that position where $\Sigma(\beta)$ has β. The phenomenon of opacity will be discussed in greater detail in Chapter 2.)

I believe that Quine first used 'opaque' to describe "constructions" for which "substitutivity" fails; roughly, a "construction" is opaque just in case coreferring terms cannot be interchanged in the construction *salva veritate*. (See Quine 1960, Chapter IV.) This way of speaking is inappropriate for my purposes because it begs the questions in which I am interested; e.g. "Does it follow from a sincere understanding subject's dissent from an occurrence of Σ, that he does not hold the attitude of belief toward the content presented by this occurrence?" And, "Does if follow from the subject's dissent from an occurrence of Σ that many occurrences of ⌜N believes that Σ⌝ (where N refers to the subject) are false?"

Second, some might object that cognitive values, or senses, were originally posited to account for "Frege's puzzle of identity" rather than the phenomenon of opacity. I maintain, however, that instances of Frege's puzzle of identity are merely instances of the phenomenon of opacity: To say that occurrences of ⌜$\alpha=\beta$⌝ "contain extensions of our knowledge," or "are not obvious," while occurrence of ⌜$\alpha=\alpha$⌝ "do not contain extensions of our knowledge," or "are obvious" (where α and β are coreferential) is to say that one can imagine a normal understanding subject who would assent to an occurrence of ⌜$\alpha=\alpha$⌝ but dissent from an occurrence of ⌜$\alpha=\beta$⌝. (Or, to put the point another way, a general explanation of the phenomenon of opacity will *a fortiori* serve as a solution to Frege's puzzle of identity.)

16 Putnam argues that narrow psychological state determines cognitive value by appealing to the Fregean conception of what it is to know the meaning of a term. Putnam argues that

there cannot be two different logically possible worlds L1 and L2 such
that, say, Oscar is in the same psychological state (in the narrow sense) in
L1 and L2 (in all respects) but in L1 Oscar understands [N] as [expressing
the cognitive value cv] and in L2 Oscar understands [N] as [expressing the
cognitive value cv']. (For, if there were, then in L1 Oscar would be in the
psychological state knowing that [cv] is the meaning of [N], and in L2
Oscar would be in the psychological state knowing that [cv'] is the meaning
of [N], and these are different, and even--assuming that [N] has just one
meaning for Oscar in each world--incompatible psychological states in the
narrow sense), [Putnam 1975, p. 221].

I think that this argument begs the question against the Fregean .theorist who is willing to
deny that narrow psychological state determines grasped cognitive value: it seems to me
that this Fregean theorist must be willing to deny the claim that knowing that cv is the
meaning of N and knowing that cv' is the meaning of N are, or involve, different
psychological states in the narrow sense. What Putnam should have argued, and what I do
argue, is that if the defender of the Fregean conception concedes that narrow
psychological state does not determine grasped cognitive value, then there is no reason to
posit cognitive values--they no longer do any explanatory work.

[17] Davidson, "On Saying That", in Davidson 1990, p. 108.

[18] Strictly speaking Frege had no truck with *propositions*, as I use the term. Frege
maintained that the cognitive value expressed by an occurrence of a sentence is a *thought*,
while the semantic value presented by an occurrence of a sentence is its *truth value*. But
even for Frege the truth value assigned to an occurrence of a sentence is determined by the
semantic values presented by the terms appearing in the sentence, and the syntactic
structure of the sentence. Hence there is no harm in interpreting Frege's theory as
involving propositions, as I use the term.

[19] The Principle of Semantic Innocence can be stated more formally, albeit less lucidly,
as follows:

> For all contexts of utterance c, declarative sentences Σ, names N,
> propositional attitude verbs ς, and Russellian propositions p, Σ presents p
> relative to c if and only if, relative to c, Σ as it appears in $\ulcorner$N ςs that $\Sigma\urcorner$
> presents p.

[20] One might object to this argument against the Fregean analysis by denying the
assumption that Mary's utterance of 'he' is strictly coreferential with John's utterance of
'Twain'. It might be suggested that even though the occurrence of 'Twain' refers to a
cognitive value, the fact that the term 'Twain' has Twain, the man, as its ordinary semantic
value suffices to make Twain salient in the context of John and Mary's conversation. Thus
'he', as uttered by Mary, can be interpreted as a "pronoun of laziness" which refers to the

most appropriate salient entity, viz. Twain. Consider, however, the following conversation:

> (e)   John: Odile believes that Twain is a great author.
>   Mary: I grasp *it* when I think of Twain too.

If 'Twain' as it occurs in John's utterance refers to a cognitive value, then this cognitive value is salient in the context of the conversation--more salient than is Twain--and thus we ought to have no difficulty in interpreting 'it' as it occurs in Mary's utterance as referring to this cognitive value. That is, if 'Twain' in John's utterance refers to a cognitive value, then (e) should be more natural than (a). Example (e), however, is not more natural than example (a); no ordinary speaker will interpret 'it' and 'Twain' as they appear in an occurrence of (e) as being coreferential.

21 Similar problems arise in examples involving factive propositional attitude verbs. For instance, there is a sense in which someone who utters an occurrence of 'Odile knows that Twain is a great author' asserts both that Odile knows that Twain is a great author, and that Twain is a great author. Frege's response to these difficulties is to posit the kind of referential ambiguity discussed above. Frege wrote that in these cases the relevant clause ". . . is to be taken twice over, with different reference, standing once for a thought, once for truth value," (Frege, "On Sense and Reference," in Frege 1966, p. 76.)

22 "Theories of Meaning and Learnable Languages," in Davidson 1990, pp. 14-5.

23 In Chapter 9 of Dummett 1973, Dummett maintains that a commitment to an infinite hierarchy of cognitive values will lead to "a reductio ad absurdum of the whole theory," (p. 267). Dummett thus attempts to amend Frege's analysis of attitude ascription so that there are referents and ordinary senses only: under Dummett's proposal, a referring term appearing inside a that-clause both *presents* (refers to) and *expresses* its ordinary sense. (Dummett concedes that in making this amendment he must deny that referents are always determined by senses.) Dummett, however, overlooks the fact that secondary senses, ternary senses, etc. will be needed by the Fregean theorist in order to account for situations such as the following: Suppose that Odile attaches the same sense to occurrences of both 'Noam' and 'Chomsky'. Odile now utters occurrences of

> (i) I believe that Noam is great thinker.

and

> (ii) I believe that Chomsky is a great thinker.

Under Dummett's proposal, the that-clauses of (i) and (ii) present the same thought, and, assuming the appearances of 'I believe that' express the same sense, the occurrences of (i) and (ii) express the same thought. Now suppose, however, that Oscar assents to Odile's utterance of (i), yet dissents from Odile's utterance of (ii). In order to account for Oscar's

behavior, a Fregean theorist must posit a difference in the senses expressed by the occurrences of (i) and (ii), but unless secondary senses are invoked, there can be no such difference.

24 Church, A. 1951.

25 The only argument Davidson offers against this "protean" interpretation of the Fregean analysis is suggested in his rather cryptic comment that "the problem is not how the individual expressions that make up a sentence governed by believes', given the [semantic values] they have in such a context, combine to denote a proposition; the problem is rather to state the rule that gives each the [semantic value] it does have." I think Davidson must have in mind something like the argument I state.

26 Frege readily admitted that the same term may express distinct senses for different subjects. (See, again, the passage concerning 'Aristotle' discussed in note 11.) It will be argued in section 2.2 that Frege is compelled to make this concession, though doing so undermines his theory.

27 One might attempt to analyze (9) as follows:

(9*) ¬∃x ∃y (y is a way of thinking of Twain &
Believes <x, y ^ CV('played quarterback')>)

(Let 'CV( )' designate a partial function from expressions to the cognitive values of occurrences of those expressions--ignoring for the time being that the cognitive value expressed by terms, and even occurrences of terms, varies intersubjectively. And let '^' denote concatenation for cognitive values; i.e. ^ is a partial function from "simple" cognitive values to "complex" cognitive values.) Our untutored intuitions dictate that an occurrence of (9) would be true, even if somebody assented to occurrences of 'Clemens played quarterback.' Yet if there were such a person, (9*) would be false. Therefore (9*) cannot be a correct analysis of (9).

28 A proper statement of the Principle of Full Articulation might require a third clause concerning the time of the occurrence. This third clause might be required to account for the fact that the same sentence can have different truth conditions if uttered at different times; e.g. an occurrence of 'Odile is hungry' at t1 might be true, while an occurrence of this same sentence at t2 might be false. This third clause can be omitted, however, if we allow the properties designated by predicates to be indexed to times; e.g. we could allow that 'is hungry' designates different properties at different times. I am not sure which of these approaches is to be preferred. Again, temporal issues will be ignored here.

29 The Naive Russellian Theory is described in greater detail by Nathon Salmon in Salmon 1986, pp. 16-18.

30 I will use capital letters to demarcate used sentences and formulas, and follow the usual practice of using numerals to demarcate mentioned sentences and formulas.

31 The "semantic similarity" clause is needed to account for cases involving indexicals, and other context sensitive terms. E.g. if Odile assents to an occurrence of 'He is a great author' in a context where 'he' refers to Twain, then it does not follow that Oscar's utterance of 'Odile believes that he is a great author' is true in context where 'he' refers to, say, Nixon.

32 This argument assumes that c is *semantically similar* to itself with regard to (3) and (4). The argument assumes, for example, that, relative to c, 'Odile' has one semantic value only.

2.0 The Fundamental Difficulty with the Fregean Strategy.

Frege's theory of sense and reference is specifically designed to account for the phenomenon of opacity; i.e. senses and thoughts are posited specifically to preclude the epistemological and semantic arguments from opacity and also to account for the cognitive significance of occurrences. The purpose of this chapter is, first, to illustrate how Frege's theory can be generalized into what I call *the Fregean strategy for precluding the arguments from opacity*, and second, to argue that any theory which utilizes this strategy will face a fundamental difficulty. The argument against Fregean theories presented in this chapter differs from the arguments presented in Chapter 1 in that the fundamental difficulty with the Fregean strategy is wholly independent of the principles of direct reference and semantic innocence; thus in this chapter I argue that there are compelling reasons for rejecting Fregean theories which are independent of the principles of direct reference and semantic innocence. (I will call theories which utilize the Fregean strategy *Fregean theories*. This term is somewhat misleading because many of the theories which I would call Fregean involve a kind of psychologism that Frege himself would find objectionable.)

The chapter proceeds as follows: In section 2.1 I briefly discuss Frege's theory of sense and reference, and describe the Fregean strategy for precluding the arguments from opacity. In section 2.2 I present what I call *the Fregean strategy's Scylla*: Though it is usually assumed that only Russellian theories are susceptible to the arguments from opacity, in fact even Fregean theories are susceptible to the arguments; if all the premises of the epistemological and semantic arguments from opacity are granted, then there is no reason to suppose that Fregean theories cannot also be refuted by arguments from opacity.[1] Moreover, in order to preclude the arguments from opacity, the cognitive values posited by Fregean theories must be individuated extremely finely, so finely that what cognitive value is presented by an occurrence of a referring term will *vary*

intersubjectively: a single occurrence of a referring term may express distinct cognitive

values for distinct subjects. But this difficulty, the difficulty of individuating cognitive

values finely enough to preclude the arguments from opacity, is only the Scylla of the

difficulty facing the Fregean strategy, and in section 2.3 I present what I call *the Fregean*

strategy's Charybdis: Because the Fregean theorist is forced to individuate cognitive

values extremely finely, the Fregean theorist undermines his ability to account for, first,

the *publicity* of the content of our propositional attitudes, and second, the legitimacy of

our ordinary attitude ascribing practices. If cognitive values--the objects of our

propositional attitudes--are individuated so finely that the cognitive value complex (or

thought) expressed by an occurrence of a sentence varies intersubjectively, then the

Fregean theorist will be unable to account for the facts that beliefs, desires etc., are

shared, and that ordinary subjects are able to ascribe such attitudes to one another.

Finally, in section 2.4 I summarize the argument against the Fregean strategy presented

in the previous sections, and I briefly discuss the significance of the fundamental

difficulty with the Fregean strategy.

2.1 The Fregean Strategy for Avoiding the Arguments and Problems Posed by the Phenomenon of Opacity.

How does Frege's theory of sense and reference endeavor to preclude the

epistemological and semantic arguments from opacity, and account for the cognitive

significance of occurrences? I begin with a brief explanation of how Frege's theory

precludes the epistemological arguments from opacity. Because Frege does not endorse

the Principle of Direct Reference, he is not compelled to maintain that the token of

'Twain' which appears in (O1) and the token of 'Clemens' which appears in (O2) express

the same content. Rather Frege posits distinct senses to serve as the contents expressed

by these tokens, and this allows him to distinguish the content expressed by (O1) from

the content expressed by (O2). More specifically, Frege posits *thoughts* to serve as the contents expressed by occurrences of declarative sentences. Thoughts are composed of senses: the thought expressed by an occurrence of a declarative sentence is composed of the senses expressed by the terms appearing in the sentence. Because the tokens of 'Twain' and 'Clemens' appearing in (O1) and w(O2) express distinct senses, it follows that (O1) and (O2) express distinct contents, or thoughts, and Frege thereby avoids the contradictory result that Odile both holds and does not hold the same attitude toward the same content. Thus, by positing senses, which are individuated more finely than referents, and maintaining that the content expressed by an occurrence of a declarative sentence--a thought--is composed of these finely individuated entities, Frege is seemingly able to preclude the epistemological arguments from opacity.

Frege's strategy for precluding the epistemological arguments from opacity can be stated slightly more formally as follows: Let us call the Fregean thought expressed by (O1) 'T', and the Fregean thought presented by (O2) 'T*'. According to Frege's theory of sense and reference it follows from Odile's assent to (O1) at t and the Epistemological Principle of Assent that

(C) Believes <Odile, T> at t.

and it follows from Odile's dissent from (O2) at t' and the Epistemological Principle of Dissent that

(D) ¬(Believes <Odile, T*> at t').

Because T and T* are not composed of identical senses, T is not identical to T* and therefore (C) and (D) are not contradictories. (More precisely, even assuming that Odile does not "change her mind" in the interval between t and t', no contradiction can be derived from (C) and (D).)

As was briefly discussed in section 1.1, the sense expressed by an occurrence of a term also accounts for the *cognitive significance* of occurrences of sentences containing the term. For example, the reason that occurrences of

(11) Twain is identical to Clemens.

seem to "contain . . . valuable extensions of our knowledge,"[2] while occurrences of

(12) Twain is identical to Twain.

do not seem to contain valuable extensions of our knowledge, is that occurrences of 'Twain' and 'Clemens' express distinct senses, or cognitive values. We judge that occurrences of (11) convey informative truths, while occurrences of (12) convey obvious truths, *because* of the senses expressed by the terms appearing in the occurrences. Similarly, according to Frege's theory of sense and reference, Odile judges that (O1) is true and thus assents to it, yet does not judge that (O2) is true and thus dissents from it *because* these occurrences express distinct thoughts: Odile assents to (O1) because she holds the attitude of belief toward the content expressed by this occurrence, and she dissents from (O2) because she does not hold the attitude of belief toward the content expressed by this occurrence. Moreover, (O1) and (O2) express distinct thoughts because the tokens of 'Twain' and 'Clemens' appearing in (O1) and (O2) express distinct senses. Thus Frege's theory of sense and reference accounts for the cognitive

significance of an occurrence by appeal to the cognitive values expressed by the terms appearing in the occurrence.[3]

How does Frege's theory of sense and reference endeavor to preclude the *semantic* arguments from opacity? Frege does not endorse the Principle of Semantic Innocence, and thus he is not compelled to maintain that (O1) presents the same proposition as the that-clause of an occurrence of

(3) Odile believes that Twain is a great author.

More precisely, because Frege does not endorse the Principle of Semantic Innocence, he is not compelled to maintain that 'Twain' as it appears in (O1) refers to the same referent as it does in an occurrence (3). Frege's famous move here is to identify the *semantic value* of a term appearing inside a that-clause with the *cognitive value* the term expresses when it appears outside of a that-clause; i.e. cognitive values serve as the *oblique referents* of referring terms. Thus according to Frege's theory, the *semantic value* of 'Twain' as it appears inside the that-clause of an occurrence of (3) is identified with the *cognitive value* which is expressed by 'Twain' as it appears in (O1). And Frege identifies the *semantic value* presented by 'Clemens' as it appears in occurrences of

(4) Odile believes that Clemens is a great author.

with the *cognitive value* which is expressed by 'Clemens' as it appears in (O2). By distinguishing the *semantic values* presented by 'Twain' and 'Clemens' when they appear in occurrences of (3) and (4), Frege is able to distinguish the propositions presented by these occurrences.[4] Consequently, according to Frege's theory of sense and reference it does not follow from the fact that an occurrence of (3) is true in a context c, that an occurrence of (4) must also be true in c. Thus, by denying the Principle of Semantic

Innocence, and alloying terms occurring inside that-clauses to have their oblique

referents as semantic values, Frege's theory of sense and reference is able to maintain that

occurrences of

(4n) Odile does not believe that Clemens is a great author.

are true, while occurrences of (4) are false.

Frege's strategy for avoiding the arguments and problems posed by the

phenomenon of opacity is thus relatively straightforward: for every referring term one

posits a number of *cognitive values*, which are something like "ways of thinking" of the

referent of the term. These cognitive values perform three central functions: First,

cognitive values serve as constituents of *content*--they are constituents of the objects

toward which subjects hold propositional attitudes. In order to serve this first function,

cognitive values must be *objective* in the sense that the same--numerically identical--

cognitive value can be accessed, or "grasped," by different subjects at different times: if,

as Frege thought, cognitive values are abstract objects, then the same cognitive value can

be grasped by different subjects at different times, and if, as many contemporary Fregean

theorists suppose, cognitive values are psychological states of some kind, then the same

cognitive value can be *instantiated* by different subjects at different times. It is this first

function of cognitive values which is responsible for precluding the epistemological

arguments from opacity: Cognitive values can be individuated much more finely than

referents, and thus the Fregean theorist hopes to avoid the contradictory conclusions of

the epistemological arguments from opacity by appropriately individuating the contents

expressed by occurrences of sentences.

The second function served by cognitive values is to account for the *cognitive*

significance of occurrences. According to a Fregean theory, Odile assents to (O1) yet

dissents from (O2) *because* 'Twain' as it appears in (O1) and 'Clemens' is it appears in (O2) express distinct cognitive values for Odile.

Finally, the third function of cognitive values is to serve was the *semantic values*, the *oblique referents*, of terms occurring inside the that-clauses of attitude ascriptions. This third function is responsible for precluding the semantic arguments from opacity: the Fregean theorist hopes to avoid the contradictory conclusions of the semantic arguments from opacity by appropriately individuating the contents presented by occurrences of the that-clauses of attitude ascriptions.

Any theory which posits *some* kind of entity to serve the above three functions-- whether these entities are called "senses," "states," "guises," "cognitive roles," "modes of apprehension," "modes of presentation," "sentences of mentalese" or what-have-you-- utilizes what I call the Fregean strategy for precluding the arguments from opacity.[5] The Fregean strategy can be more precisely defined in terms of Ramsey sentences; i.e. we can identify the class of Fregean theories as the class of theories which endorse, or entail, a certain set of Ramsey sentences. To this end, let us first define *an instance of the phenomenon of opacity* as a situation involving a subject and occurrences of two sentences $\Sigma(\alpha)$ and $\Sigma(\beta)$ which meets the following necessary and jointly sufficient conditions:

> (1) The subject (who is sincere, normal, and non reticent) is in some
> manner presented with a token of $\Sigma(\alpha)$ and a token of $\Sigma(\beta)$, where α and β
> are coreferrential; e.g. the subject hears utterances of, or sees or feels
> inscriptions of, $\Sigma(\alpha)$ and $\Sigma(\beta)$. ('$\Sigma(\ )$' is a variable ranging over open
> sentences, and 'α', and 'β' are variables which range over proper names.
> Thus '$\Sigma(\alpha)$' and '$\Sigma(\beta)$' are open for replacement by quotation names for
> sentences which differ only in that in one position they have distinct, yet
> coreferring, referring terms.)

(2) The subject fully understands occurrences of both the token of $\Sigma(\alpha)$ and the token of $\Sigma(\beta)$. (I assume that it is compatible with a subject's "fully understanding" tokens of $\Sigma(\alpha)$ and $\Sigma(\beta)$ that the subject not know that α and β are coreferential.)[6]

(3) The subject assents to the token of $\Sigma(\alpha)$, yet dissents from the token of $\Sigma(\beta)$.

Fregean theories can now be characterized as the set of theories T which satisfy the following schema:

> Theory T posits a domain of cognitive values, and/or provides a principle of individuation for cognitive values, and for all instances of the phenomenon of opacity O involving a subject named N and occurrences of $\Sigma(\alpha)$ and $\Sigma(\beta)$, T maintains that a proper account of O will entail the following Ramsey sentence (where the domain of the quantifiers is the posited domain of cognitive values):
>
> $\exists w \, \exists x \, \exists y \, \exists z$ [(w is expressed by $\Sigma(\alpha)$) & (x is expressed by $\Sigma(\beta)$) & (y is expressed by α) & (z is expressed by w) & (w contains y as a constituent) & (x contains z as a constituent) & (($y \neq z$) and therefore ($w \neq x$)) &
>
> (i) (x and y are contents--they are the kind of thing that serve as the objects of propositional attitudes) &
>
> (ii) (the subject assents to $\Sigma(\alpha)$ *because* $\Sigma(\alpha)$ expresses w and the subject holds the attitude of belief toward w, while the subject dissents from $\Sigma(\beta)$ *because* $\Sigma(\beta)$ expresses z and the subject does not hold the attitude of belief toward z) &
>
> (iii) (the that-clauses of relevant occurrences of ⌜N believes that $\Sigma(\alpha)$⌝ present w as semantic value) & (the that-clauses of relevant occurrences of ⌜N believes that $\Sigma(\beta)$⌝ present x as semantic value)]

(Note that a "complete" Ramsey sentence would contain no theoretical terms, and thus would be much more complicated than the above sentence; a complete Ramsey sentence would replace appearances of theoretical terms such as 'presents', 'semantic value' 'express' etc. with either non theoretical terms, or additional bound variables.[7])

Fregean theories are thus committed to there being a domain of entities--cognitive values of some kind--which serve the three functions cognitive values were posited to serve. I argue in the following sections, however, that there cannot be any such entities. The fundamental difficulty with the Fregean strategy is that there are can be no entities which serve all three of the functions cognitive values were posited to serve.

2.2 The Fregean Strategy's Scylla.

The arguments from opacity against the Naive Russellian Theory require that there be (or that there *could* be) an instance of the phenomenon of opacity; they require that there be a subject such as Odile who assents to an occurrence such as (O1), yet dissents from an occurrence such as (O2). Let us say that Odile's situation, as described in Chapter 1, is *an instance of opacity with regard to R*, where R is the Naive Russellian Theory according to which (O1) and (O2) express the same content; i.e. according to R (O1) and (O2) present the same Russellian proposition. More generally, let *an instance of opacity with regard to theory T* be an instance of opacity such that the occurrences of $\Sigma(\alpha)$ and $\Sigma(\beta)$ are said by T to *express the same content*. (Recall that 'content' is a theory neutral term designating the objects of our propositional attitudes, whatever these may be.) Consequently, assuming the requisite principle of compositionality of content, there is *an instance of opacity with regard to theory T* if and only if there is an instance of opacity involving occurrences of $\Sigma(\alpha)$ and $\Sigma(\beta)$, and according to T the tokens of α and β which appear in these occurrences express the same content.

It is at least in principle possible for there to be instances of opacity with regard to theories which utilize the Fregean strategy. For example, suppose that Fregean theory F employs a principle of individuation for cognitive values according to which the cognitive value presented by occurrences of both 'Cicero' and 'Tully', for a subject Oscar, is identified with the content presented by the description 'a famous Roman orator'. (Of

59

course no actual theorist is likely to hold such a undiscriminating principle of individuation; Frege certainly did not.[8]) If theory F maintains that occurrences of both 'Tully' and 'Cicero' express this one cognitive value for Oscar, then according to F occurrences of both

(13) Cicero orated.

and

(14) Tully orated.

express the same content for Oscar. (Let us call this content 'C'.) Now let us suppose that our subject Oscar thinks that there are two famous Roman orators--one named 'Cicero' and the other named 'Tully'--and as a result Oscar assents to an occurrence of (13), yet dissents from an occurrence of (14). (I will call these occurrences '(O13)' and '(O14)' respectively.) Since simple minded theory F maintains that (O13) and (O14) express the same content for Oscar, Oscar's situation is *an instance of opacity with regard to F*.

Moreover, and this is the crux of the point being made here, there being an instance of opacity with regard to F is necessary and sufficient for there being both epistemological and semantic arguments from opacity against F. Continuing with the above example, it follows from Oscar's assent to (O13) at t and the Epistemological Principle of Assent that

(E) Believes <Oscar, C> at t.

And it follows from Oscar's dissent from (O14) at t' and the Epistemological Principle of Dissent that

(F) ¬(Believes <Oscar, C>) at t'.

Therefore, assuming that Oscar does not "change his mind" in the interval between t and t', epistemological arguments from opacity against the simple minded Fregean theory can be formulated.

The above example involving Oscar and (O13) and (O14) can also be utilized to formulate semantic arguments from opacity against Fregean theory F. It follows from Oscar's assent to (O13) and the Semantic Principle of Assent that in some context c, an occurrence of

(15) Oscar believes that Cicero orated.

is true. And it follows from Oscar's dissent from (O14) and the Semantic Principle of Dissent that in c

(16n) Oscar does not believe that Tully orated.

is also true. (For the sake of simplicity, let us assume that there is *one* context c in which both (15) and (16n) are true.) According to simple minded theory F, however, for Oscar occurrences of 'Tully' express the same cognitive value as occurrences of 'Cicero', and consequently in any context in which (15) is true, an occurrence of

(16) Oscar believes that Tully orated.

61

will also be true. Since (15) is true in c, (16) is also true in c. But it was previously determined that (16n), which is simply the negation of (16), is true in c. Since a sentence and its negation cannot both be true in the same context, the argument concludes that the simple minded theory F must be rejected.

The defender of the Fregean strategy can, of course, accept the above arguments as refutations of simple minded Fregean theory F, without rejecting the Fregean strategy in general. Theory F is, after all, very simplistic; it is surely implausible to suppose that for Oscar the tokens of 'Tully' and 'Cicero' appearing in (O13) and (O14) have as cognitive value the content expressed by 'a famous Roman orator'. The above arguments do illustrate, however, that it is possible to construct arguments from opacity against theories which utilize the Fregean strategy; contrary to what is often assumed, merely positing some kind of cognitive values to serve as the objects of our propositional attitudes and the semantic values of that-clauses is not in itself sufficient for precluding the arguments from opacity. Rather, in order to preclude the arguments from opacity, the cognitive values posited by a Fregean theory must be appropriately individuated.[9]

Armed with the terminology introduced above, a general statement of conditions necessary and sufficient for constructing arguments from opacity against any semantic theory can be formulated:

> The Principle of Arguments from Opacity: For all theories T, there is a nomologically possible *instance of opacity with regard to T* iff corresponding epistemological and semantic arguments from opacity against T can be formulated.

Note that the principle does not state, nor does it entail, that if there is an instance of opacity with regard to T, then T can be *refuted*. The principle merely states that if there is an instance of opacity with regard to T, then there are corresponding epistemological and semantic arguments from opacity against T, but one is not compelled to accept these

arguments as refutations of T unless one grants all of the premises of the arguments.[10] Also note the modal operator: in order to formulate an argument from opacity against T, one need not produce an *actual* instance of opacity with regard to T; it is sufficient to show that there *could be* an instance of opacity with regard to T. The relevant modal notion here is not *logical* possibility. We seek an explanation of the phenomenon of opacity as it involves human beings in the actual world, and consequently we are interested only in those possible instances of opacity that involve beings similar to us in worlds similar to ours. Therefore the relevant modal notion is *nomological possibility*: it is nomologically possible that p iff, roughly, there is a possible world w in which the laws of physics, biology, psychology, etc., are similar to the laws which obtain in the actual world, and p is true in w. (Or, if one objects to possible worlds talk, it is nomologically possible that p iff given the laws of nature as they actually obtain, p could be true.)

It follows from the above principle that there being epistemological and semantic arguments from opacity against a Fregean theory T is contingent upon there being a nomologically possible instance of opacity with regard to theory T. Furthermore, there being a nomologically possible instance of opacity with regard to theory T is contingent upon the particular account of cognitive values proffered by theory T. The above argument against simple minded Fregean theory F can be constructed because F does not distinguish the cognitive values expressed by the tokens of 'Cicero' and 'Tully' appearing in (O13) and (O14) for Oscar, though Oscar assents to (O13) yet dissents from (O14). This illustrates that if a Fregean theory T is to preclude the arguments from opacity against it, then T must provide an account of cognitive value such that there are no nomologically possible instances of opacity with regard to T. In other words, in order to preclude the arguments from opacity, a Fregean theory T employing the Fregean strategy must satisfy the following constraint:

The Individuation Constraint: If there is a nomologically possible
instance of opacity involving a subject A, and occurrences of sentences
$\Sigma(\alpha)$ and $\Sigma(\beta)$, then T must distinguish the cognitive values expressed by
these occurrences of α and β for A.

Or equivalently,

It must be *nomologically impossible* for there to be an instance of opacity
with regard to T.

What is often overlooked is that for any Fregean theory T, it is an open question as to

whether theory T satisfies the individuation constraint.

Satisfying the individuation constraint is hardly a trivial task. For a Fregean

theorist to claim that his theory T satisfies the individuation constraint is for him to claim

that *for every nomologically possible instance of opacity involving a subject A and*

occurrences of referring terms α and β, according to T, these occurrences express

distinct cognitive values for A. But how are we to understand this strong modal claim,

and how can it be justified? There are two ways the Fregean theorist's strong modal

claim can be interpreted.

First, the Fregean theorist's strong modal claim could be interpreted as a *semantic*

claim concerning the meaning of the term 'cognitive value' (or 'sense' or what-have-you).

If the Fregean theorist's claim is interpreted in this way, it is simply *definitive* of the

cognitive values posited by T that for any nomologically possible instance of opacity

involving subject A, and occurrences of α and β, these occurrences of α and β present

distinct cognitive values for A. Consequently, under this semantic interpretation of the

Fregean theorist's strong modal claim, sentences such as the following are analytic:

'Odile assents to (O1), yet dissents from (O2) only if the tokens of 'Twain' and 'Clemens'

appearing in these occurrences present distinct cognitive values for Odile.' (Or at least

they are analytic if anything is.) Therefore, if the Fregean theorist's claim is interpreted

as a semantic claim, then his theory T trivially satisfies the individuation constraint: Under this first interpretation it is nomologically impossible for there to be an instance of opacity with regard to theory T in the same way that it is nomologically impossible for there to be a married bachelor.

The problem with interpreting the Fregean theorist's strong modal claim as a semantic claim is that it precludes theories which utilize the Fregean strategy from providing *explanations* of the phenomenon of opacity; i.e. it precludes such theories from accounting for the *cognitive significance* of occurrences. If theory T simply defines 'cognitive value' (or what-have-you) so that the individuation constraint is necessarily satisfied, then asserting that "Odile assents to (O1), yet dissents from (O2), *because* 'Twain' and 'Clemens' express distinct cognitive values for Odile" is like asserting that "Oscar is not married because he is a bachelor." Even if these sentences are true, and arguably they are not, they certainly do not provide us with acceptable *explanations*. One of the functions cognitive values are posited to serve is to account for the cognitive significance of occurrences; cognitive values are to explain why Odile assents to (O1), yet dissents from (O2), despite the fact that (O1) and (O2) in some sense say the same thing about the same person. If the Fregean theorist's strong modal claim is interpreted semantically, if it is true merely by fiat, then the Fregean theory cannot hope to provide such explanations.

The Fregean theorist's strong modal claim must, therefore, be interpreted in a second way; viz. it must be interpreted as a *metaphysical* claim concerning the nature of a certain kind of entity. The Fregean theorist must be understood as positing cognitive values as independent entities--entities whose identity conditions can be stated without appeal to the phenomenon of opacity. If the Fregean theorist's claim is interpreted in this second way, then Fregean theories can, at least in principle, provide an acceptable, informative explanation of the cognitive significance of occurrences: Odile assents to (O1), yet dissents from (O2), *because* the tokens of 'Twain' and 'Clemens' appearing in

these occurrences express distinct cognitive values for Odile, and these cognitive values have intrinsic properties which, somehow, account for Odile's verbal behavior. (Or perhaps it would be more appropriate to say that these cognitive values are *instantiated* by entities, perhaps token brain states, which have intrinsic properties which account for Odile's verbal behavior.) However, if cognitive values are understood as such independent entities allegedly responsible for the phenomenon of opacity, then the Fregean theorist is committed to there being, at least in principle, a means of individuating these entities which is both *principled* and *independent*. If Fregean theory T is to provide an adequate explanation of the phenomenon of opacity, then the Fregean theorist who advocates T is committed to there being general principles which govern how cognitive values are individuated; there must be some fact of the matter that makes cognitive value x distinct from cognitive value y. Furthermore, these general principles must involve *independent* individuation conditions; i.e. principles which do not themselves appeal to the phenomenon of opacity.

The situation is analogous to Moliere's example concerning the soporific effects of opium: the phenomenon to be explained is that when a normal subject, say Odile, imbibes opium, she falls asleep. We might attempt to explain this phenomenon by positing a certain property, "dormitivity," and claiming that Odile's falling asleep is due to the fact that opium is a "dormitive substance."[11] But if we are now asked, "What is it for a substance to have the property of dormitivity? How is the property of dormitivity distinguished from other properties?" we cannot, on pain of undermining our explanation, claim that for a substance to have the property of dormitivity just is for the substance to have soporific effects on people who imbibe it. That is, in attempting to say what it is for a substance to have the property of dormitivity we cannot, on pain of undermining our explanation, appeal to the very phenomenon the property was posited to explain. Analogously, cognitive values are posited to account for the phenomenon of opacity, and consequently if they are to be explanatorily relevant to this phenomenon, it

must be at least in principle possible to individuate them wholly independently of this phenomenon. (That Odile assents to an occurrence of $\Sigma(\alpha)$, yet dissents from an occurrence of $\Sigma(\beta)$, cannot be constitutive of the occurrences of α and β expressing distinct cognitive values for Odile; something other than Odile's behavior must make it the case that the occurrences of a and b express distinct cognitive values for Odile.)[12]

Moreover, if the Fregean theorist violates the Independence Constraint (see below) and thereby undermines the explanatory relevance of cognitive values with regard to the cognitive significance of occurrences, his strategy for precluding the epistemological and semantic arguments from opacity is also rendered *ad hoc*. Concerning the epistemological arguments, if cognitive values were irrelevant to the explanation of Odile's assent to (O1) and dissent from (O2), then there would be no motivation *independent of the epistemological arguments from opacity* for the claim that 'Twain' and 'Clemens' as they appear in (O1) and (O2) present distinct cognitive values. If cognitive values could not be appealed to in order to explain Odile's judgment and behavior with regard to (O1) and (O2), then there would be no motivation for supposing that the objects of Odile's attitudes were cognitive value complexes beyond the fact that in positing cognitive values to serve as contents, the Fregean theorist precludes the epistemological arguments from opacity. More generally, if cognitive values were irrelevant to the explanation of cognitive significance, there would be no reason for supposing that cognitive values serve as the constituents of contents, and the Fregean strategy's method of precluding the epistemological arguments from opacity would be rendered *ad hoc*. And concerning the semantic arguments from opacity, if there were no good reason for supposing that cognitive values served as the constituents of contents, then why suppose that cognitive values serve as the oblique referents of terms appearing inside the that-clauses of attitude ascriptions? For the Fregean theorist is surely correct in assuming that the semantic values of the that-clauses of attitude ascriptions are contents; the semantic values of that-clauses are the objects of our propositional attitudes.

67

Thus if the only reasons for supposing that cognitive values serve as contents are *ad hoc*, it follows that the reasons for supposing that cognitive values serve as oblique referents are also ad hoc. the upshot is that if the Fregean theorist violates the Independence Constraint, not only does he forfeit the explanatory relevance of cognitive values to the problem of the cognitive significance of occurrences, but his strategy for precluding the epistemological and semantic arguments from opacity is also rendered *ad hoc*.

If the above is correct, then in order to preclude the arguments from opacity, and also provide an account of the cognitive significance of occurrences, a Fregean theory must satisfy both of the following metaphysical constraints: First, in order to preclude the arguments from opacity, a Fregean theories T must satisfy

> The Individuation Constraint: For every nomologically possible instance of opacity involving a subject A, and occurrences of sentences $\Sigma(\alpha)$ and $\Sigma(\beta)$, T distinguishes the cognitive values expressed by these occurrences of α and β for A.

and, second, in order to account for the cognitive significance of occurrences, and thereby provide an explanation of the phenomenon of opacity, a Fregean theory T must satisfy

> The Independence Constraint: The cognitive values posited by T are individuated wholly independently of the phenomenon of opacity; it is in principle possible to state the individuation conditions for the cognitive values posited by T without appealing to the phenomenon of opacity.[13]

The challenge posed by the Fregean strategy's Scylla is the challenge of providing an account of cognitive values which satisfies both of the above constraints. I maintain that this challenge has yet to be met by a Fregean theory; i.e. every proposed account of cognitive value I know of fails to satisfy at least one of the above constraints. (In Appendix A I examine some proposed accounts of cognitive value, and show that each of

these proposals fails to satisfy one or the other of the above constraints.) My central argument against the Fregean strategy, however, does not presume that a Fregean theorist cannot provide an account of cognitive value which satisfies the above constraints, nor even that a Fregean theorist has yet to do so. Rather my central argument depends only upon the much weaker thesis that a Fregean theory can satisfy the above constraints only if it individuates cognitive values extremely finely, so finely that what cognitive value is presented by an occurrence of a referring term will *vary intersubjectively*. That is, a Fregean theory can satisfy the above two constraints (the individuation constraint in particular) only if it allows occurrences of referring terms to present distinct cognitive values for distinct subjects. I now turn to supporting this weaker thesis.

It is clear that if a Fregean theorist is to have any hope at all of satisfying the individuation constraint, he must be capable of individuating cognitive values at least as finely as the expression types (of the tokens) which present them; i.e. the Fregean must be at least capable of individuating cognitive values such that for all subjects X and distinct term types α and β, occurrences of α and β present distinct cognitive values for X. That the Fregean must be capable of individuating cognitive values at least as finely as expression types is obvious: for any pair of coreferring terms α and β which are not of the same type, it is *nomologically possible* for there to be an instance of opacity involving α and β. (For any two coreferring term types α and β, it is nomologically possible for there to be a sincere, understanding subject like Odile who understands occurrences of these terms, yet does not believe that they are coreferential. Moreover, if it is nomologically possible for there to be such a subject, then it is nomologically possible for there to be some $\Sigma(\ )$ such that the subject assents to an occurrence of $\Sigma(\alpha)$, yet dissents from an occurrence of $\Sigma(\beta)$.) Thus the Fregean theorist might attempt to satisfy the individuation constraint and the independence constraint by utilizing the following principle of individuation for cognitive values: "No two occurrences of distinct expression types can present the same cognitive value, for any subject."

It is at least plausible to suppose that a Fregean theory utilizing such a principle of individuation could satisfy the individuation constraint and the independence constraint. But, first, it should be noted how counterintuitive the proposal is; the claim that (tokens of) no two expression types can express the same cognitive value, where cognitive values are the objects of our propositional attitudes, is both drastic and *ad hoc*. What is left of our Fregean intuitions if occurrences of 'steed' and 'horse' cannot express the same cognitive value, or sense? And what about our intuitions concerning foreign languages? Surely one who holds the attitude of belief toward the content expressed by an occurrence of 'Schnie ist weiss' also holds the attitude of belief toward the content expressed by an occurrence of 'Snow is white'. The Fregean theorist might try to preserve our intuitions by invoking a translation relation between expressions of different languages. The Fregean could then form equivalence classes of expression types under this relation, and claim that terms within the same equivalence class express the same cognitive value. Notice, however, that the translation relation would have to do more than preserve reference: the Chinese term for 'Twain', for example, could not be a translation of 'Clemens'. But given this kind of difficulty, how is the translation relation to be determined without appeal to the phenomenon of opacity?

Second, and more importantly, though it is certainly *necessary* for the Fregean to be able to distinguish cognitive values as finely as expression types, it is not sufficient to individuate cognitive values only as finely as expression types. Even this somewhat drastic principle for individuating cognitive value does not individuate cognitive values finely enough to satisfy the individuation constraint. As Kripke's "Paderewski" example illustrates, there are nomologically possible instances of opacity in which the relevant occurrences of coreferential terms α and β are instances of the same term type; i.e. there are nomologically possible situations in which a sincere, understanding subject falsely thinks that different occurrences of the same referential term refer to different things.[14] If Peter thinks that there is a famous pianist named 'Paderewski', and that there is a

famous statesman named 'Paderewski', and Peter also believes that statesmen never have musical talent, then there may be some context c in which Peter dissents from an occurrence of

(17) Paderewski is a musician.

and another context c' in which Peter assents to an occurrence of this same sentence. (There is nothing all that special about Kripke's "Paderewski" example. The phenomenon is common in cases involving indexicals and demonstratives: of course a subject might assent to one occurrence of 'he is a musician' and dissent from another occurrence of 'he is a musician', where the two occurrences of 'he' refer to the same person.) According to a Fregean theory which individuates cognitive values only as finely expression types, these occurrences would express the same content, and consequently, by the Principle of Arguments from Opacity, it would be possible to formulate both epistemological and semantic arguments from opacity against this Fregean theory. Merely individuating cognitive values according to the expression types (of the tokens which express the cognitive values) does not individuate cognitive values finely enough; something more is needed.

It is apparent that what is needed is a principle of individuation for the cognitive values presented by referring terms which appeals to something internal to the subjects who use the term. The problem with individuating cognitive values by appeal to expression types only is that the expression type of an occurrence used or perceived by a subject is not directly related to the internal state of the subject and thus is not directly relevant to explaining the subject's judgments and behavior. Thus the Fregean must claim that differences in cognitive values are determined by differences in "the ways" subjects understand occurrences of expressions, where "ways of thinking" are some kind of mental properties, states, or abilities of the subject. But what, exactly, is a "way of

thinking"? There are, of course, many proposals: One might appeal to "modes of apprehension" where a the mode of apprehension of a term is something like the *kind of experience* the subject underwent when he learned how to use the term. (For example, the mode of apprehension expressed by occurrences of 'Twain' for Odile might be some kind mental image of a man in white suit--a memory of a picture she was looking at when she learned the name 'Twain'.[15]) Or perhaps the cognitive value expressed by an occurrence of a term is an *individual concept* the subject associates with the occurrence. Every property P determines an individual concept: a property P* such that x has P* just in case x i⌐ the unique thing that has P. (For example, Odile may think of Twain via the individual concept, the man who wrote *The Adventures of Huckleberry Finn*.[16]) Another proposal is to appeal to the "functional role" the term (or the "mentalese correlate" of the term) plays in the subject's mental life. Under this proposal occurrences of 'Twain' and 'Clemens' express distinct cognitive values for Odile just in case these terms (or their "mentalese correlates") play different *roles* in Odile's inference patterns, attitudes, and behavior. (For example, occurrences of 'Twain' and 'Clemens' express distinct cognitive values for Odile if Odile is disposed to infer from 'Twain is witty' to 'The author of *The Adventures of Huckleberry Finn* is witty', but is not disposed to infer from 'Clemens is witty' to 'The author of *The Adventures of Huckleberry Finn* is witty'.[17])

I maintain that none of these proposals can succeed in satisfying the individuation constraint and the independence constraint, and in Appendix A I provide some support for this strong thesis. Here, however, I am concerned only to defend the weaker thesis that in order to individuate cognitive values finely enough to satisfy the individuation constraint, the cognitive value expressed by an occurrence of a referring term must be permitted to *vary intersubjectively.* If I am correct in maintaining that Paderewski-like cases compel the Fregean theorist to invoke mentalistic principles of individuation for cognitive values such as those listed above, then this weaker thesis must be correct. Our mental lives are extremely different from one another: distinct subjects may, and often

do, have different educational backgrounds, different experiences, different beliefs and desires, and even, dare I say, different *Weltanschauungen*. If such detailed features of our mental lives are invoked in order to individuate cognitive values, then the cognitive values expressed by occurrences of referring terms will vary from subject to subject. More specifically, distinct occurrences of the same referring term might express different cognitive values relative to the same subject; one occurrence might express distinct cognitive values for distinct subjects; and distinct occurrences of the same term might express distinct cognitive values for distinct subjects. Whether or not the cognitive value(s) presented by an occurrence of a referring term varies intersubjectively in one of these ways is contingent upon "the way" the subject(s) think(s) about the referent of the term, where this "way of thinking" is determined by some detailed feature of the subject's mental state.

2.3.1 The Fregean Strategy's Charybdis.

If the Fregean theorist is to have any hope of steering clear of Scylla, the problem of providing an account of cognitive values which satisfies the individuation and independence constraints, he must permit the cognitive values presented by occurrences of terms to vary intersubjectively. However, if the Fregean theorist permits the cognitive values expressed by occurrences to vary intersubjectively, he steers directly toward Charybdis, the problems of accounting for the publicity of our propositional attitudes, and preserving the legitimacy of our ordinary propositional attitude ascribing practices. I will consider these problems in turn.

2.3.2 The Problem of Preserving the Publicity of Content.

Frege maintained a sharp distinction between *thoughts*, the contents of our propositional attitudes, and *ideas*, the subjective phenomenological experiences of consciousness. Thoughts, which are composed of senses, are objective, public, entities capable of being grasped by different individuals at different times. Ideas, on the other hand, are subjective, private, entities which "need an owner" in order to exist.[18] Thus the fundamental difference between thoughts and ideas is that thoughts are essentially *communicable*, while ideas are essentially *incommunicable*. Dummett explains

> Frege was the first to attach due weight to the fact that we cannot have a thought which we do not express, to ourselves if not to others. . . .
> Thought differs from other things also said to be objects of the mind, for instance pains or mental images, in not being essentially private. I can tell you what my pain is like, or what I am visualizing, but I cannot transfer to you my pain or my mental image. It is the essence of thought, however, that it is transferable, that I can convey to you exactly what I am thinking . . . I do more than tell you what my thought is like--I communicate to you that very thought.[19]

The problem of accounting for the publicity of content can be stated as follows: if cognitive values, the constituents of thoughts, or contents, are individuated by mentalistic criteria such as those briefly discussed above, then contents cannot serve as that which is *communicated* by occurrences.

Suppose, for example, that Odile and Oscar are apparently having a disagreement concerning Twain. Odile is impressed with Twain's wit, and thus she is led to utter an occurrence of

(18) Twain is witty.

Oscar, on the other hand, is not amused by Twain's anecdotes, and thus he is led to respond by uttering an occurrence of

(18n) Twain is not witty.

If, however, occurrences of 'Twain' express distinct cognitive values for Odile and Oscar, then, despite appearances, there is no incompatibility between what Odile and Oscar believe--if Odile and Oscar "think of Twain in different ways," then there is nothing in the exchange which implies that they stand in the believes relation to contradictory contents. If 'Twain' as it appears in Odile's utterance has as its cognitive value one finely individuated "way of thinking of Twain," and 'Twain' as it appears in Oscar's utterance has as its cognitive value a distinct "way of thinking of Twain," then there is nothing incompatible in the contents expressed by Odile and Oscar's utterances. Odile and Oscar are, we might say, "expressing contents at cross purposes."

Similarly, if cognitive values are permitted to vary intersubjectively, cases of apparent agreement between two subjects will not in fact be cases of agreement. Suppose, for example, that Odile admires Twain's writing style and is thus led to utter an occurrence of

(1) Twain is a great author.

And suppose that Oscar is also fond of Twain's writing style, and thus he responds by nodding and uttering a distinct occurrence of (1). Again, Odile and Oscar's utterances of (1) *present* the very same proposition, and thus their utterances have the same truth conditions. But, again, if they utilize distinct "ways of thinking of Twain," then they do not *believe* the same thing--their utterances express distinct contents, and thus despite

75

appearances Odile and Oscar are not manifesting agreement in what they believe. Odile and Oscar are, again, "expressing contents at cross purposes."

Examples such as the above illustrate that one of the consequences of permitting cognitive values to vary intersubjectively is a widespread *incommensurability of content*. If cognitive values, the constituents of content, are permitted to vary intersubjectively, then cases in which conversing subjects hold propositional attitudes toward the same content will be the exception, and not the rule. Some philosophers, including Frege and Dummett, have thought that this kind of *incommensurability of content* would undermine our ability to communicate. If communication between A and B requires that A and B "grasp," or instantiate, the very same content, yet A and B attach distinct cognitive values to occurrences of referring terms, then it is of course going to be extremely difficult for A and B to succeed in communicating. Frege was well aware of this problem; in a letter to Jourdain Frege wrote,

> Now if the sense of a name was something subjective, then the sense of
> the [sentence] in which the name occurs, and hence the thought, would
> also be something subjective, and the thought that one man connects with
> this [sentence] would be different from the thought another man connects
> with it; a common store of thoughts, a common science would be
> impossible. It would be impossible for something one man said to
> contradict what another man said, because the two would not express the
> same thought at all, but each his own.[20]

Frege hoped to avoid the problems posed by the incommensurability of content by maintaining a sharp distinction between objective *senses* and *thoughts* on the one hand, and subjective *ideas* on the other. If, however, Fregean theorists are forced to individuate cognitive values by appeal to detailed psychological factors, then Frege's distinction between *objective* senses, and *subjective* ideas becomes blurred, and widespread incommensurability of content ensues.

One might attempt to accommodate the incommensurability of content by denying the Fregean model of communication; i.e. one might attempt to avoid the result that communication is undermined by denying that A and B communicate only if A and B "grasp," or instantiate the very same content. One could maintain, for example, that communication between A and B requires only that A and B think *of* the same proposition, as opposed to "thinking of the same proposition in the same way"; i.e. perhaps communication between A and B requires only that A and B think *of* the same referents and denotations, and does not require that they think of these referents and denotations "in the same ways." Or perhaps communication between A and B merely requires that A and B think *of* the same proposition in similar enough ways. I am, however, skeptical about these proposals. It seems to me that these proposals encroach upon the essential communicability of content. The Fregean model of communication ought to serve as a constraint upon conceptions of content: contents, the objects of our propositional attitudes, are whatever is shared and communicated by competent speakers of a language. If under a certain conception of content it turns out that Odile and Oscar, in the above example, are not expressing and believing the same content, then one must reject that conception of content. In other words, the Fregean conception of communication is correct: the contents of our propositional attitudes are, and must be, what is communicated, and if the Fregean conception of content cannot preserve this identity, then so much the worse for the Fregean conception of content. These complex and murky issues, however, are somewhat beyond the scope of my present concerns, and thus I will not pursue them here. Rather here I will be concerned to illustrate that even if the Fregean theorist is able to account for communication by denying the Fregean model of communication, the incommensurability of content raises other problems which are not so easily avoided.

Recall that cognitive values are posited to serve not only was the contents expressed by occurrences of referring terms, but also as the semantic values of those

terms when they appear inside the that-clauses of occurrences of attitude ascriptions. Permitting cognitive values to vary intersubjectively thus results not only in incommensurability of content, but also results in a kind of incommensurability of semantic value, or reference. This incommensurability of semantic value undermines the Fregean theorist's ability to account for the validity of simple arguments whose premises contain attitude ascriptions. Consider, for example, the following argument:

1. Odile believes that Twain is witty.
2. Oscar does not believe that Twain is witty.
Therefore,
3. Odile believes something that Oscar does not believe.

This argument is intuitively valid. It seems to be an instance of the valid argument schema

1. R<a, p>
2. ¬(R<b, p>)
Therefore,
3. ∃x (R<a, x> & ¬(R<b, x>))

If cognitive values are permitted to vary intersubjectively, however, the above argument is not an instance of this schema. The problem is that if cognitive values are permitted to vary intersubjectively, the that-clauses of premises 1 and 2 cannot be interpreted as two occurrences of the same singular term: if cognitive values are permitted to vary intersubjectively, the appearances of 'Twain' in these premises may *refer to* distinct "ways of thinking of Twain," and as a result the that-clauses of premises 1 and 2 may have distinct thoughts as semantic values. Thus, if cognitive values are permitted to vary intersubjectively, the existential generalization by way of which the conclusion is derived is not warranted. This result, however, is unacceptable. The existential generalization is warranted; ordinary speakers make such inferences all the time.

It is important to see that this kind of difficulty is not limited to a few special cases; if cognitive values are allowed to vary intersubjectively, almost all intuitively valid arguments involving propositional attitude ascriptions will turn out to be invalid. Consider the intuitively valid argument

1. Odile believes everything that Oscar believes.
2. Oscar believes that Twain is witty.
Therefore,
3. Odile believes that Twain is witty.

The problem is, again, that if cognitive values are allowed to vary intersubjectively, then it does not follow from the form of the argument that the that-clauses in premises 2 have the same thought as semantic value. Thus, if cognitive values are permitted to vary intersubjectively, this argument is not an instance of the valid schema

1. $\forall x\, (R\langle a, x\rangle \rightarrow R\langle b, x\rangle)$
2. $R\langle a, p\rangle$
Therefore,
3. $R\langle b, p\rangle$

Again, however, this result is unacceptable. We commonly employ such arguments, and we intuit that they are valid. An adequate analysis of the logical form of these arguments must respect these untutored intuitions.[21]

I conclude that if cognitive values are permitted to vary intersubjectively, a Fregean theorist will be unable to account for the publicity of content: if the Scylla of the Fregean strategy forces the Fregean theorist to permit the cognitive values presented by occurrences to vary intersubjectively, then the Fregean theorist will be unable to account for the communicability of content, and he will be unable to account for our untutored intuitions concerning the validity of arguments involving attitude ascriptions.

2.3.3 The Problem of Preserving the Legitimacy of Our Ordinary attitude Ascribing Practices.

According to a Fregean theory, an occurrence of a belief ascription is true if and only if the subject of the ascription holds the attitude of belief toward the content presented by the that-clause of the ascription. Furthermore, the content presented by the that-clause of an ascription is a thought--it has cognitive values as constituents. Once cognitive values are permitted to vary intersubjectively, however, it is far from clear that this can be a proper analysis of propositional attitude ascriptions. Consider, for example, Oscar's utterance of

(3) Odile believes that Twain is a great author.

According to the Fregean analysis of attitude ascriptions, an occurrence of (3) is true just in case Odile stands in the believes relation to the thought presented by the that-clause of the occurrence. But, in order to individuate cognitive values finely enough to avoid the arguments from opacity, the Fregean must claim that what is presented by the that-clause in an occurrence of (3) varies intersubjectively; i.e. it is quite likely that Oscar and Odile "think of Twain in different ways," and consequently it is quite likely that occurrences of the that-clause of (3) present distinct thoughts for Oscar and Odile. if one occurrence of a that-clause presents distinct thoughts for two subjects, then how, in the absence of special knowledge concerning one another's "ways of thinking of referents," can one subject truly ascribe an attitude to the other? I assume that our ordinary attitude ascribing practices are legitimate: Ordinary subjects are capable of making true attitude ascriptions concerning the beliefs, desires, doubts etc. of other subjects, and they are capable of doing so even in the absence of special knowledge concerning "the ways" the subjects of their ascriptions think of things. But how can the Fregean analysis of attitude

ascriptions preserve the legitimacy of our ordinary attitude ascribing practices if cognitive values are permitted to vary intersubjectively?

Consider again Oscar's utterance of the ascription (3). What thought is presented by the that-clause of Oscar's utterance of (3)? There are only two proposals worthy of consideration: the *speaker oriented proposal,* and the *subject oriented proposal.* The speaker oriented proposal maintains that the thought presented by the that-clause of an attitude ascription is determined by the *person who utters the ascription*; in terms of the above example, the cognitive values Oscar assigns to the terms appearing in the that-clause determine the thought presented by the that-clause. The subject oriented proposal, on the other hand, maintains that the thought presented by the that-clause of an ascription is determined by the *subject of the ascription*; in terms of the above example, the cognitive values Odile assigns (or would assign) to the terms appearing in the that-clause determine the thought presented by the that-clause. I will argue that neither proposal is compatible with the legitimacy of our ordinary attitude ascribing practices.[22]

Suppose that the Fregean theorist opts for the speaker oriented proposal; i.e. suppose that an occurrence of a belief ascription is true if and only if the subject of the ascription holds the attitude of belief toward the content that the speaker assigns to the that-clause of his utterance. I will present three objections to this proposal.

First, this proposal is susceptible to refutation via the semantic arguments from opacity. Suppose that Oscar assigns the same cognitive values to occurrences of both 'Twain' and 'Clemens'. (Suppose he acquired the names when he was told that "there is a famous author who is called both 'Twain' and 'Clemens'".) Now consider, again, the instance of opacity in which Odile assents to an occurrence of

(1) Twain is a great author.

yet dissents from an occurrence of

(2) Clemens is a great author.

(I will continue to call these occurrences '(O1)' and '(O2)' respectively.) Now suppose that, in a context c, Oscar sincerely utters occurrences of both

(3) Odile believes that Twain a great author.

and

(4n) Odile does not believe that Clemens is a great author.

It follows from the semantic principles of assent and dissent that Oscar's utterances of (3) and (4n) are true. Oscar, however, assigns the same cognitive value to occurrences of both 'Twain' and 'Clemens', and consequently, according to the speaker oriented proposal, Oscar's utterance of (3) in c is true if and only if an utterance, by Oscar, of (4) Odile believes that Clemens is a great author. is true in c. But, again, (4n) is simply the negation of (4), and a sentence and its negation cannot both be true relative to the same context.

The above argument against the speaker oriented view assumes that it is nomologically possible for a speaker to assign the same cognitive values to occurrences of distinct terms α and β, and it also assumes that it is possible for such a subject to utter true ascriptions of the form $\ulcorner N$ believes that $\Sigma(\alpha)\urcorner$ and $\ulcorner N$ believes that $\neg\Sigma(\beta)\urcorner$, where N refers to the subject. Thus the Fregean theorist opting for the speaker oriented proposal might try to avoid this objection by denying either one of these assumptions. Denying either assumption, however, would be counterintuitive and *ad hoc*. First, the Fregean theorist must grant that it is nomologically possible for a subject to assign the same

cognitive value to distinct coreferring terms. To deny this assumption would be to deny that a bi-lingual speaker can assign the same cognitive value to distinct terms, but this is implausible and *ad hoc*. What independent reason could there be for supposing that it is nomologically impossible for a bi-lingual speaker of French and English to assign the same content to occurrences of 'London' and 'Londres'? Furthermore, it is extremely implausible to suppose that speakers *must* assign different cognitive values to different name variations: surely those who know James Higginbotham well assign the same content to occurrences of both 'James Higginbotham' and 'Jim Higginbotham'. Concerning the second assumption, to claim that Oscar cannot truly utter an occurrence of (3) and (4n) because he assigns the same "way of thinking" of Twain to occurrences of both 'Twain' and 'Clemens' is simply to abandon our ordinary attitude ascribing practices: if Odile assents to (O1), yet dissents from (02), then, other things being equal, Oscar's utterances of (3) and (4n) are true, regardless of the "way(s) in which Oscar thinks of Twain."[23]

My second objection to the speaker oriented proposal concerns what Richard has called the *Echo Principle*.[24] Suppose Odile swears up and down that she believes that Twain is a great author; suppose even that she is renowned for sincerely uttering occurrences of both

(1) Twain is a great author.

and

(19) I believe that Twain is a great author.

Our untutored intuitions dictate that if Oscar utters an occurrence of

(3) Odile believes that Twain is great author.

and his justification for making this utterance is Odile's sincere utterances of (1) and (19),

then Oscar's utterance of (3) is true. More generally, our ordinary attitude ascribing

practices respect the following principle:

> The Echo Principle: If a normal, sincere, understanding subject A named
> N utters an occurrence of Σ, or an occurrence of $\ulcorner$I believe that $\Sigma\urcorner$ in a
> context c, and if *on the basis of this utterance* another normal,
> understanding subject B utters an occurrence of $\ulcorner$N believes that $\Sigma\urcorner$ in a
> semantically similar context c', then, *ceteris paribus*, B's utterance is
> true.[25]

My second objection to the speaker oriented proposal is that it is incompatible with the

Echo Principle.

Suppose that Oscar and Odile assign distinct cognitive values to occurrences of

'Twain'. Further suppose, what is quite likely the case, that Odile does not hold the

attitude of belief toward any content which has as a constituent the cognitive value Oscar

assigns to occurrences of 'Twain'. And finally suppose that on the basis of Odile's sincere

utterances of (1) and (18), Oscar is led to utter an occurrence of

(3) Odile believes that Twain is a great author.

Assuming that Odile has not revised her beliefs, our untutored intuitions, encapsulated in

the above statement of the Echo Principle, dictate that Oscar's utterance of (3) is true.

According to the speaker oriented proposal, however, Oscar's utterance of (3) is false:

despite Odile's sincere utterances of (1) and (19), if occurrences of 'Twain' present

distinct cognitive values for Odile and Oscar, and Odile does not hold the attitude of

belief toward any content which has as a constituent the cognitive value Oscar assigns to

occurrences of 'Twain', then Oscar's utterance of (3) is false. If cognitive values are permitted to vary intersubjectively, then it is quite likely that the that-clause in Oscar's occurrence of (3) does not express a thought toward which Odile holds the attitude of belief. This result, however, is unacceptable. If Odile swears up and down that she believes that Twain is a great author, and if on the basis of such utterances Oscar utters an occurrence of (3) then, other things being equal, our untutored intuitions dictate that Oscar's utterance of (3) is true.

My third and final objection to the speaker oriented proposal is that it is incompatible with another principle which is respected by our ordinary attitude ascribing practices. If Odile is renowned for sincerely uttering occurrences of (1) and (18), then in semantically similar contexts occurrences of

(3n) Odile *does not* believe that Twain is a great author.

are false. More generally, our ordinary attitude ascribing practices respect the following "Dis-Echo Principle":

> The Dis-Echo Principle: If a normal, sincere, understanding subject A named N utters an occurrence of Σ, or an occurrence of $\ulcorner$I believe that $\Sigma\urcorner$, in a context c, and another normal understanding subject B utters an occurrence of $\ulcorner$N *does not* believe that $\Sigma\urcorner$ in a semantically similar context c', then, *ceteris paribus*, B's utterance is false.

The situation described above will also serve to illustrate that the speaker oriented proposal is incompatible with the Dis-Echo Principle. If, as is assumed in the above situation, Odile does not hold the attitude of belief toward any content which contains the cognitive value Oscar assigns to occurrences of 'Twain', then, despite Odile's sincere utterances of (1) and (19), the speaker oriented proposal will predict that Oscar's

(relevant) utterance of (3n) will be *true*. Again, however, this result is unacceptable; if Odile sincerely utters occurrences of (1) and (18), then Oscar's utterance of (3n), in a semantically similar context, is false.

If the Fregean theorist adopts the speaker oriented proposal--which maintains that the thought expressed by a that-clause within an ascription is determined by *the person who utters the ascription*--then semantic arguments from opacity against his theory will be forthcoming, and he will be unable to preserve the untutored intuitions which are encapsulated in the Echo and Dis-Echo Principles. I conclude that the speaker oriented proposal does not preserve the legitimacy of our ordinary attitude ascribing practices, and therefore it must be rejected. I now turn to the subject oriented proposal.

According to the subject oriented proposal, the thought presented by the that-clause of an occurrence of an attitude ascription is determined by the cognitive values the *subject of the ascription* assigns (or would assign) to the terms appearing in the that-clause of the ascription. In terms of the above example, the subject oriented proposal maintains that the thought presented by the that-clause of Oscar's utterance of (3) is determined by the cognitive values Odile assigns to the terms appearing in that-clause. (Of course an immediate problem arises because there may not be, and probably is not, a one-to-one correspondence between term types and cognitive values; e.g. Odile probably assigns distinct cognitive values to different occurrences of 'Twain'. Let us, for the time being, overlook this difficulty.) My argument against the subject oriented proposal proceeds as follows: I first present a number of more-or-less technical difficulties with the proposal, and I consider, and reject, various ways of amending the proposal to avoid these technical difficulties. I conclude with a more general, conceptual, objection which applies to any version of the subject oriented proposal.

The subject oriented proposal, as it is stated above, cannot account for cases in which the subject of the ascription is not familiar with all of the terms appearing in the that-clause of the ascription. Suppose, for example, that a Chinese speaker, Mao,

believes that Twain is a great author, though he assigns no cognitive value whatsoever to utterances or inscriptions of 'Twain'. (Suppose that Mao assents to occurrences of the Chinese translation of (1).) In such a situation our untutored intuitions dictate that Oscar's utterance of

(20) Mao believes that Twain is a great author.

would be true. The subject oriented proposal predicts, however, that Oscar's utterance of (20) will be neither true nor false: Since Mao assigns no cognitive values whatever to occurrences of 'Twain' the subject oriented proposal predicts that 'Twain' as it appears in the that-clause of Oscar's utterance has no referent, and thus according the subject oriented proposal Oscar's utterance of (20) suffers from reference failure.[26]

The obvious move for the Fregean to make here is to appeal to some kind of translation relation between English and Chinese. More specifically, the Fregean theorist might attempt to avoid the above difficulty by defining a translation relation R, which holds between terms of all languages. He could then form equivalence classes of terms under R, and modify the subject oriented proposal was follows:

> The Modified Subject Oriented Proposal: The cognitive value referred to by a referring term α appearing the that-clause of an attitude ascription is the cognitive value the subject of the ascription assigns to occurrences of any term α' which is a member of C, where C is the equivalence class of α under R.

There are several difficulties with this proposal. First, this proposal succeeds in avoiding the above difficulty only if there are no two terms α and α', both of which are members of the equivalence class formed by 'Twain' under R, such that Mao assigns distinct cognitive values to occurrences of α and α'. But how can the translation relation be

specified so as to ensure that this cannot occur, without appealing to the phenomenon of opacity? Second, it is not at all clear how this proposal can be extended to cover cases involving demonstratives and indexicals. For example, how could the modified subject oriented proposal be extended to account for the intuitive truth of *Twain's* utterance of

(21) Mao believes that I am a great author.

If Oscar's utterance of (20) is true, then, other things being equal, Twain's utterance of (21) is also true. Thus the modified subject oriented proposal must claim that 'I'--at least as it appears in Twain's utterance of (20)--is a translation of Mao's name for Twain. Consequently, if the modified theory is to be extended to cover ascriptions involving indexicals and demonstratives, the translation relation must hold between *occurrences* of terms. But how can a translation relation which holds between occurrences be defined *beforehand*? How can we translate future occurrences of 'I'? The modified subject oriented proposal can account for the truth of Twain's utterance of (21) only if 'I' as it appears in that occurrence is a translation of Mao's name for Twain. But how could we have known before Twain uttered (21) that *this* occurrence of 'I' was a translation of Mao's name for Twain? (A similar difficulty is presented by cases in which the only term *the subject employs* is a demonstrative. For example, suppose that Mao has never heard of Twain under a name, but he comes to believe that Twain is a great author by, say, watching him write a book--Mao, we could say, has a "demonstrative belief" about Twain that he is a great author. If this situation obtains, then there are many contexts in which wan utterance of (20) would be true, yet it is not at all clear how the modified subject oriented proposal can account for the truth of these occurrences.)

The Fregean theorist might attempt to avoid the above difficulties by appeal to yet another version of the subject oriented proposal. The difficulties presented above arise because Mao, the subject, is not appropriately related to a *term* which appears in the

that-clause of the speaker's ascription. It seems that these problems would not arise if one could state what cognitive value is referred to by a referring term appearing in a that-clause without appealing the term itself. Thus the defender of the subject oriented proposal might suggest the following *referential* version of the proposal:

> The Referential Version of the Subject Oriented Proposal: The cognitive value referred to by a referring term appearing in the that-clause of an ascription is "the way (or perhaps *a* way) in which the subject thinks of the *referent* of the term."

According to the referential version of the subject oriented proposal, Oscar's utterance of

(20) Mao believes that Twain is a great author.

expresses the proposition that Mao holds the attitude of belief toward a content which has Mao's own "way of thinking of Twain" as a constituent, regardless of whether Mao assigns this cognitive value to occurrences of 'Twain', or translations of 'Twain'.

Again, however, there are several difficulties with this version of the subject oriented proposal. First, it does not specify what cognitive value is presented by a referring term appearing inside of a that-clause, because, assuming that such a term has a referent, it is unlikely that the subject of the ascription has only one way of thinking of that referent. When Oscar utters an occurrence of

(3) Odile believes that Twain is a great author.

to which one of Odile's "ways of thinking" of Twain is Oscar referring? More importantly, the proposal seems to abandon the spirit of the Fregean approach. The Fregean approach, as I understand it, is an internalist, Cartesian, view of content and

propositional attitudes; it maintains that our beliefs, desires, etc. can to be specified and individuated independently of the actual state of "the external world." But in specifying the contents expressed by occurrences of referring terms the referential version of the subject oriented proposal makes an essential appeal to the referents of those terms, and in doing so the proposal abandons the internalistic spirit of the Fregean approach.

Not surprisingly, abandoning the spirit of the Fregean approach in favor of the referential version of the subject oriented proposal has a price. One of the alleged advantages of Fregean theories concerns their ability to account for our untutored intuitions with regard to occurrences of ascriptions such as

(22) Odile believes that Santa is fat.

There are situations in which our untutored intuitions dictate that an occurrence of (22) is true. The Fregean theorist who does not endorse the referential version of the subject oriented view can account for these intuitions because the content expressed by the that-clause of occurrences of (22) in no way depends upon 'Santa' having an ordinary referent. If, however, the referential version of the subject oriented proposal is adopted, this virtue of Fregean theories will be lost: there obviously is no "way in which Odile thinks of the referent of 'Santa'." (This point illustrates the tension between Frege's claim that "empty" referring terms present senses, on the one hand, and his account of senses as "modes of presentation of *objects*" on the other. What is a "mode of presentation" of nothing?[27])

Another difficulty, shared by all versions of the subject oriented proposal, is that it cannot account for our intuitions concerning the truth conditions of occurrences of ascriptions such was

(8) Aristotle believed that Twain was a great author.

The term 'Twain' as it appears in occurrences of the above ascription of course has a referent, and thus it at least *makes sense* to speak of "Aristotle's way of thinking of Twain." In this case the difficulty is not that there is no referent, but rather that there is no such "way of thinking" of the referent: Aristotle did not think of Twain in any "way" at all. Thus even the referential version of the subject oriented proposal will incorrectly predict that occurrences of (8) are neither true nor false.

A similar difficulty is presented by occurrences of ascriptions such as

(23) Nobody believes that Twain is a great author.

and

(24) Everybody believes that Twain is a great author.

The problem presented by occurrences of (23) and (24) is not that the subject has no "way of thinking of Twain," rather the problem is that there is no particular subject whose "way of thinking of Twain" could serve as the semantic value of 'Twain'.[28] All versions of the subject oriented proposal will predict that no particular content is expressed by occurrences of such ascriptions, but, intuitively, this prediction is not correct. our untutored intuitions suggest that an occurrence of (23) asserts that nobody holds the attitude of belief toward what is expressed by the that-clause of the occurrence, where the that-clause of the occurrence serves to *specify* the relevant content. Note that there is no straightforward way of solving this problem by appealing to "hidden" quantification over cognitive values. For example, (24) cannot plausibly be analyzed was having the following logical form:

(24*) $\forall x \exists y$ (y is a "way of thinking of Twain" &
Believes<x, y^CV('is a great author')>).

91

(Let 'CV( )' designate a function from expression types to their cognitive values--for convenience I assume that the cognitive values of predicates do not vary intersubjectively. And let '^' designate a concatenation device for cognitive values; more precisely, '^' designates a partial function from "simple" cognitive values to "complex" cognitive values and contents.) The reason that occurrences of (24) cannot be analyzed along the lines of (24*) is that (24*) would be true even if everybody *dissented* from occurrences of

(1) Twain is a great author.

yet assented to occurrences of

(2) Clemens is a great author.

The fact that everybody assented to occurrences of (2) would suffice to make an occurrence of (24*) true, but our untutored intuitions dictate that in such a situation an occurrence of (24) would be *false*. Consequently nothing along the lines of (24*) can serve as a Fregean analysis of (24). (Similar remarks apply to w(23), and other ascriptions which fail to specify a particular subject.)

I conclude my arguments against the subject oriented proposal with a more general, conceptual, objection which applies to all versions of the proposal. Simply put, the objection is that the subject oriented proposal is not a plausible analysis of what ordinary speakers do when they utter attitude ascriptions. In saying of someone that he "fears that it's raining," or "wonders whether Jeeves did it," or "believes that the Sox will take it this year," does an ordinary speaker really refer to certain finely individuated "ways of thinking"? It certainly does not seem as if in making ordinary attitude

ascriptions we are referring to such entities; rather it seems that in uttering an occurrence of 'Oscar believes that the Sox will take it this year,' one is simply describing the world as Oscar believes it to be, viz. such that the Sox are going to take it this year. The that-clauses of ordinary attitude ascriptions do not seem to specify "ways of thinking" of the world; rather they seem to refer to conceivable *states of the world*. My final objection to the subject oriented proposal is that things are not as they do not seem to be: it cannot be that in making attitude ascriptions ordinary speakers refer to finely individuated "ways of thinking," because there are occurrences of true (and false) attitude ascriptions where the speaker simply lacks the ability to refer to an appropriate "way of thinking."

Consider Oscar's utterance of

(3) Odile believes that Twain is a great author.

Under the subject oriented proposal, if this occurrence is to be true, the occurrence of 'Twain' must have one of Odile's "ways of thinking" of Twain as its semantic value. But what determines the referent of this occurrence? What makes it the case that this particular "way of thinking" is the referent of the occurrence?[29] Though this is not the appropriate place to undertake an analysis of reference, it is, I think, generally agreed that if a subject uses a term α to refer to an object O, then there must be an explanation as to why this particular occurrence of α is associated with the particular object O. As Fregean theorists insist, *something* must determine the referent of an occurrence; something *must make it the case* that *this* is the referent of the occurrence. (Note that we are here concerned with "speaker reference," or "occurrence reference," as opposed to "semantic reference," or "type reference"; it is clear that there is no language or idiolect in which the type 'Twain' has one of Odile's cognitive values as its referent.) I know of only two kinds of proposal concerning what the reference determining mechanism might be: First, there are, broadly speaking, *descriptive* theories of reference determination.

According to such theories, an occurrence of α refers to O just in case the speaker associates with α (or this particular occurrence of α) a criterion for identifying O: The criterion may be manifested in the form of a description, or a weighted cluster of descriptions, which uniquely denotes O, or as an ability to recognize O, or an ability to distinguish O from relevant alternatives.[30] Second, there are *causal* theories of reference determination.[31] According to causal theories an occurrence of α refers to O just in case there is an appropriate "causal chain" linking this occurrence of α to O. (These kinds of theories of reference are not incompatible: one might consistently propound a mixed theory, part causal, part identity criterion.[32]) Neither of these kinds of theories can adequately explain how a term appearing in the that-clause of an ordinary attitude ascription can be used to refer to a cognitive value.

Consider Oscar's utterance of

(3) Odile believes that Twain is a great author.

Under the subject oriented proposal, if this occurrence is to be true, the occurrence of 'Twain' must have one of Odile's "ways of thinking" of Twain as its semantic value. But what determines the referent of this occurrence? What makes it the case that this particular "way of thinking" is the referent of this occurrence? It is clear that causal theories of reference can be of no avail where: there is no appropriate causal chain linking the appearance of 'Twain' in Oscar's utterance to some "way in which Odile thinks of Twain." (Even assuming that Odile utilizes such a "way of thinking," this cognitive value was certainly never "baptized" 'Twain'.) Consequently, if the Fregean theorist is to explain how Oscar's utterance of 'Twain' manages to refer to some particular "way in which Odile thinks of Twain," he must invoke some kind of descriptivist theory of reference. Since it can be safely assumed that Oscar has never perceived a cognitive value, it cannot be that Oscar's utterance of 'Twain' refers to a particular cognitive value

because he has some perceptual capacity to recognize that cognitive value, or because he has the perceptual ability to distinguish that cognitive value from other relevant cognitive values. Consequently, if Oscar's utterance of 'Twain' refers to a particular cognitive values, it must be that Oscar "mentally employs" a description, or a cluster of descriptions, which is uniquely satisfied by one of Odile cognitive values, i.e. one of her "ways of thinking of Twain." the difficulty is that ordinary speakers can utter true (or false) attitude ascriptions even if they lack the ability to formulate a description which uniquely denotes an appropriate cognitive value.

Suppose, as is likely, that Odile thinks of Twain in at least two different ways: Suppose that Odile thinks of Twain as the author of *The Adventures of Huckleberry Finn*, (Let us call this "way of thinking" 'CV1'.) and that she also thinks of Twain was the author of *Tom Sawyer*. (Let us call this second "way of thinking" 'CV2'.) Further suppose that Odile is not confused: she believes that CV1 and CV2 determine the same referent. (This may, or may not, require her to possess, or be capable of grasping, a third cognitive value which is some kind of fusion of CV1 and CV2.[33]) And finally, suppose that Oscar's utterance of (3) is intuitively true--suppose that Odile is prone to sincerely utter sentences such as (1). Which cognitive value, CV1 or CV2, has Oscar referred to in uttering (3)? If Oscar's utterance of (3) is to have a truth value, he must have mentally employed a description which denotes only one of CV1 or CV2. But Oscar may not possess the ability to formulate such a description: suppose that Oscar thinks of Twain as "the greatest American author," and he has never even heard of *The Adventures of Huckleberry Finn*, or *Tom Sawyer*. How can Oscar formulate a description which uniquely denotes CV1, as opposed to CV2, if he has no knowledge whatsoever of what distinguishes CV1 from CV2?[34] (Note that if the Fregean strategy is to succeed in precluding the arguments from opacity, Oscar's utterance of (3) cannot be interpreted as an existential claim; it cannot be analyzed as asserting that "There is some cv which determines Twain, and Odile holds the attitude of belief toward cv^CV('is a great

author')." For if occurrences of (3) are analyzed in this way, then occurrence of (4) must be given the very same analysis, and thus the Fregean theorist will be unable to distinguish the truth conditions of occurrences of (3) and (4).)

A subject might think of an entity in myriad different ways; the more a subject knows, or believes, about an object, the more "ways of thinking about the object" he will be capable of employing. (If this is right, then the difficulties presented by the intersubjective variance of cognitive values will be most acute in situations which involve, instead of famous individuals like Twain, familiar persons and objects.) Under the subject oriented proposal, if a speaker is to make a true (or false) attitude ascription concerning what that subject believes (etc.) about that entity, he must be able to specify by description one of these "ways of thinking." But there is no reason to suppose that an ordinary subject who makes a true (or false) attitude ascription will possess such an ability. Indeed, given how little ordinary speaker's know (or even believe) concerning one another's "ways of thinking," it seems that ordinary speakers will rarely possess such an ability. I conclude that the subject oriented proposal cannot be a correct account of what ordinary speakers do when they make attitude ascriptions.

This general, conceptual, objection to the subject oriented proposal concludes my objections to the subject oriented proposal, and thus also concludes my exposition of the second half of the Fregean strategy's Charybdis. If the Scylla of the Fregean strategy forces the Fregean theorist to permit the cognitive values presented by occurrences of referring terms to vary intersubjectively, then the Fregean analysis of attitude ascriptions must adopt either the speaker oriented proposal, or the subject oriented proposal. Neither proposal, however, preserves the legitimacy of our ordinary attitude ascribing practices.

2.4 A Brief Summary.

It is often alleged that Fregean theories are superior to Russellian theories because it is possible to formulate arguments from opacity against Russellian theories, while it is not possible to formulate arguments from opacity against Fregean theories. It was shown in section 2.2, however, that this claim is simply false; it is at least in principle possible to formulate semantic and epistemological arguments from opacity against Fregean theories. Furthermore, in order to preclude refutation by way of arguments from opacity, Fregean theories must individuate cognitive extremely finely, so finely that the cognitive value presented by an occurrence of a referring term will vary intersubjectively. But this difficulty, the difficulty of providing an independent principle of individuation for cognitive values which individuates them finely enough to preclude the arguments from opacity, is only the Fregean strategy's Scylla; once cognitive values are permitted to vary intersubjectively the Fregean theorist must face Charybdis, the problems of accounting for the publicity of the contents of our propositional attitudes, and preserving the legitimacy of our ordinary propositional attitude ascribing practices. I argued in section 2.3 that, once the cognitive values presented by an occurrence of a referring term is permitted to vary intersubjectively, Fregean theories are unable to provide solutions to these problems.[35] The conclusions reached in sections 2.2 and 2.3 taken together imply that the Fregean strategy suffers from a fundamental difficulty: there could not be any entities which could serve all three of the functions cognitive values were posited to serve.

The arguments of this chapter complete my case against the Fregean approach to propositional attitudes and propositional attitude ascriptions. Not only are Fregean theories objectionable won the grounds that they violate the principles of direct reference and semantic innocence, but they are also objectionable for wholly independent reasons. Fregean theories abandon what seem to be extremely plausible general semantic

principles in order to provide an acceptable account of propositional attitudes and propositional attitude ascriptions, but the account of propositional attitudes purchased at the cost of these general principles is in a way internally inconsistent, and thus is wholly unacceptable. Consequently, since the Fregean approach exacts a high price, yet yields little or no benefit, I suggest that the Fregean approach be rejected in favor of the Russellian approach, which respects the principles of direct reference and semantic innocence. Thus in the remaining chapters I will be concerned to formulate an acceptable Russellian theory; i.e. a Russellian theory which is not subject to refutation via the problems and arguments based upon the phenomenon of opacity.

Chapter 2 Notes

[1] Of course if one or the other of the arguments from opacity is unsound, then this argument poses no threat to either Russellian theories or Fregean theories.

[2] Frege, "On Sense and Reference," in Frege 1966, p. 56.

[3] Note that Odile would assent to (O1) yet dissent from (O2) only if she would assent to relevant occurrences of (12), yet dissent from relevant occurrences of (11). This reflects the fact that the essence of the phenomenon of opacity is the phenomenon of recognition failure. Frege's theory of sense and reference is first and foremost a theory of recognition failure. These issues are further explored in Appendix A.

[4] Again, strictly speaking Frege had no truck with *propositions*, as I use the term, but there is no harm in interpreting Frege's theory as involving propositions. Frege maintained that the truth value of a sentence is compositionally determined by the semantic values--the referents and designations--of the terms appearing in the sentence. I have merely added a step in this compositionality: the semantic values of the terms appearing in a sentence compositionally determine a proposition, and the proposition thus determined has the property of truth, or the property of falsity.

[5] As was explained in Chapter 1, according to Frege's conception of reference and content the cognitive value presented by an occurrence of a referring term is also responsible for *determining the referent of the occurrence*. This function of cognitive values, however, is not directly relevant to the Fregean strategy for precluding the arguments from opacity, and thus will be overlooked here.

[6] I will not undertake the daunting task of analyzing the notion of *understanding an occurrence*. Suffice it to say that understanding an occurrence requires knowing the meanings of the words which make up the occurrence, and, following Wittgenstein and others, at least a sufficient condition for knowing the meaning of a word is possessing an ability to use the word appropriately. Notice that this conception of understanding an occurrence allows for the possibility of a subject understanding occurrences of both 'Twain is a great author' and 'Clemens is a great author', even though he does not know that 'Twain' and 'Clemens' are coreferential. Ancient astronomers had false beliefs concerning 'Hesperus' and 'Phosphorus', but they were certainly competent in using these terms to refer to Venus.

[7] Ramsey sentences are first presented in Ramsey's paper, "Theories," in Braithwaite 1931. They are explained and discussed in Block, 1980, and in Lewis, "How to Define Theoretical Terms," in Lewis 1983.

[8] Frege, unfortunately, never provides identity conditions for senses. Frege does however seem to be committed to there being (at most) one referent or designation for every sense. Furthermore, Frege seems committed to the idea that senses are

epistemologically transparent in the sense that one who "grasps" a sense would not (or perhaps cannot) "take" the sense to correspond with referent r at t1, and distinct referent r' at t2. (Of course to maintain that a proper account of senses will have it that senses possess these properties is not to provide an account of senses.) Thus Frege would deny that occurrences of 'Tully' and 'Cicero' express the content expressed by occurrences of 'a famous Roman orator', for Oscar, if Oscar thought that Tully and Cicero were two people.

9 To my knowledge, Scheffler was the first to point out that theories which utilize the Fregean strategy also face the arguments from opacity. In "On Synonymy and Indirect Discourse" (1955), found in Scheffler 1955, Scheffler, first, hints at the distinction between the epistemological and the semantic arguments from opacity, and second, points out that positing cognitive values does not in itself preclude the arguments from opacity. Scheffler directs his argument against Carnap's notion of "intensional isomorphism," but the point can be adjusted to apply to the Fregean strategy in general. Scheffler writes,

> To exclude the possibility that [Odile] may truly be said to [believe that Twain is a great author but not that Clemens is a great author] is to express a psychological theory as well as a semantical one, and a highly improbable one at that. For the same limitations which prevent [Odile] from seeing that [an occurrence of] a sentence [expresses the same singular proposition] as another may prevent [her] from seeing that one sentence [expresses the same thought as] another. . . Generalizing now, if our notion of the relation between [a that-clause] and some appropriate that-content is to be faithful to presystematic use and psychologically plausible, then no matter how semantically narrow this relation is conceived, short of identity, it will be two wide for some cases.

(Of course Kripke's "Paderewski" case illustrates that even identifying "that-content" with that-clauses will not individuate "that-content" finely enough.)

Salmon also realizes that the arguments from opacity can also be used against theories which utilize the Fregean strategy. Salmon does not notice, however, that his theory employs a version of the Fregean strategy in order to avoid the epistemological arguments from opacity, and thus is susceptible to epistemological arguments from opacity. This will be investigated in section 3.1 . See Salmon 1986.

10 Here I foreshadow my response to the epistemological arguments from opacity: In section 4.1 I will argue that the *Epistemological* Principle of Dissent should be rejected, though I maintain, unlike many other Russellian theorists, that the *Semantic* Principle of Dissent cannot be rejected. Salmon, Soames, Barwise and Perry, and Richard have all, at one time or another, proposed rejecting the Semantic Principle of Dissent. See Salmon 1986; Soames 1987; Barwise and Perry 1983; and Richard, "Direct Reference and Ascriptions of Belief," in Salmon and Soames (eds.), 1988.

11 Moliere, *Le Malade Imaginaire*, act three.

12 Perhaps another analogy can clarify this last point. Consider a fictional case based upon Mendel's work in genetics. Suppose that we are interested in accounting for the phenomenon of *differing offspring flower color*. The observable facts of this phenomenon are as follows: Sometimes a red flowered plant coupled with a red flowered plant produces a red flowered plant, while other times a red flowered plant coupled with a red flowered plant produces a white flowered plant, and still other times a red flowered plant coupled with a red flowered plant produces a pink flowered plant. How can this phenomenon be explained? Similarly to the way in which a Fregean theory attempts to account for the phenomenon of opacity, we could posit entities called "geno-types" and claim that the flower color of offspring is determined by the "geno-types" of the parent plants. Our initial theory might run as follows: if both parent plants have "red geno-types," then they will produce red flowered offspring. If both parent plants have "white geno-types," then they will produce white offspring, and if one parent plant has "red geno-types" and the other "white geno-types" then they will produce pink flowered offspring. In giving this account of the phenomenon of differing offspring flower color, we commit our theory to satisfying an individuation constraint for geno-types, and we thus commit ourselves to several strong modal claims reminiscent of the Fregean theorist's strong modal claim. For example, we are committed to the claim that for every nomologically possible instance in which a white flowered plant is produced from two red flowered plants, both of the parent plants must have "white geno-types." How is this strong modal claim to be understood? Again, if our theory is to provide an explanation of the relevant phenomena, our posited geno-types must be understood as independent entities which are responsible for the phenomenon of differing offspring flower color. But if our strong modal claim is interpreted as such a metaphysical claim, we then commit ourselves to there being, at least in principle, a way of individuating geno-types which does not appeal to the phenomenon of differing offspring flower color. We commit ourselves to there being some way of individuating geno-types by appeal to say, the strands of DNA present in the cells making up the plants, or the mass of the root systems, or the acidity of the fluid in the stems, or some combination of such independent properties. We are committed to this claim because if geno-types cannot be individuated independently of the phenomenon of offspring flower color, then we cannot explain this phenomenon by appeal to these entities. We cannot say that Red Plant A and Red Plant B produced a white Plant C *because* A and B both have, or are of, white geno-types, if geno-types cannot be individuated without appealing to the phenomenon of differing offspring flower color.

[13] In Schiffer 1990, Schiffer argues in favor of a similar constraint, which he calls the *Intrinsic-Description Constraint*:

> if a thing plays the role of [a cognitive value], then it must be intrinsically identifiable in a way that does not describe it as a [cognitive value]. In other words, it must be possible to answer the question, "What is the [cognitive value] of so-and-so?" in terms of an intrinsic characterization of the [cognitive value] whose meaning implies nothing about the thing it applies to being a [cognitive value]. If a thing is a [cognitive value], it must be intrinsically identifiable as some other kind of thing.

(Schiffer uses 'mode of presentation' instead of 'cognitive value'.) Schiffer presents the following argument in support of his claim:

If there really are things which play the [role of cognitive values], then it is a contingent fact about them that they play that role. Consequently, if there are [cognitive values], then it must be possible to say what they are in terms that do better than to characterize them as [cognitive values], or as potential [cognitive values].

[14] Kripke 1979.

[15] Salmon seems to have something like this in mind when he posits "modes of acquaintance." See Salmon 1986, pp. 107-9. As will be explained in Chapter 3, Salmon's "modes of acquaintance" do not serve as the semantic values of terms appearing inside that-clauses--but of course there is a possible theory in which they do play such a role.

[16] This seems to be the essence of Frege's view. I owe this way of putting it to Schiffer 1990, and Schiffer 1989, Chapter 3.

[17] Views appealing to "conceptual," "functional", or "cognitive" roles are defended in Block 1986, and Fodor "Propositional Attitudes" in Fodor 1981, and Fodor "Substitution Arguments and the Individuation of Beliefs," in Fodor 1990. A more detailed view involving conceptual roles is developed in Boer and Lycan 1986. (Boer and Lycan's view is criticized in Richard 1990. Views appealing to conceptual roles are also discussed in Schiffer 1992, Schiffer 1990, and Schiffer 1989.

[18] Frege, "Thoughts," in Salmon and Soames, (eds.) 1988.

[19] Dummett, "Frege's Distinction Between Sense and Reference," found in Dummett 1978, pp. 116-7.

[20] Gabrial, Hermes, Kambartel, Theil and Veraart (eds.) 1980. (I found the passage in Evans 1982, p. 15.)

21 One might attempt to preserve the validity of arguments involving attitude ascriptions by claiming that the quantifiers in these arguments are *substitutional*, rather than objectual. There are several difficulties with this move.

First, because the proposition expressed by an occurrence is partly a function of the time of the occurrence, it will be extremely difficult to specify the substitution class of the quantifiers. For example, arguments such as the following are invalid:

1. Odile believes that it's raining.
2. Odile will believe tomorrow everything that she believes today.
Therefore,
3. Tomorrow Odile will believe that it's raining.

In order to ensure that such arguments are not predicted to be valid, 'it's raining' will have to be precluded from the substitution class of the quantifiers. But then the Fregean will not be able to account for intuitively valid arguments in which 'it's raining' would have to be a substitution instance. (I owe this argument to Richard 1990, p. 76-8.)

Second, and more importantly, it seems that the only motivation a *Fregean theorist* might have for appealing to substitutional quantification would be the hope that expression types could serve to individuate cognitive values. As Paderewski-like cases make clear, however, this hope is unfounded.

22 These two proposals are considered in Richard 1990, in Forbes 1990. I am indebted to both Richard's and Forbes' discussion of these issues.

23 Forbes, in Forbes 1990, brings out this problem is a slightly different way. Suppose, first, that Oscar assigns the same cognitive value to occurrences of 'Twain' wand 'Clemens', and, second, that Odile assents to occurrences of 'Twain is witty and Clemens is not witty.' According to the semantic principle of assent, it follows that relevant occurrences of

Odile believes that Twain is witty and Clemens is not witty.

are true. But how can Oscar report Odile's belief that Twain is witty and Clemens is not witty? If Oscar were to utter an occurrence of the above sentence, he would be asserting that Odile holds the attitude of belief toward an obviously contradictory content: Let 'cv' refer to the cognitive value Oscar assigns to occurrences of both 'Twain' and 'Clemens', and let 'CV(<_,_>)' designate a (partial) function from ordered pairs ⸱ f individuals and expressions to the cognitive values the individual assigns to occurrences of those expressions. Furthermore, let '^' denote concatenation for cognitive values. (Thus, ^ is a partial function from "simple" cognitive values to "complex" cognitive values.) Allowing for a few shortcuts,* Oscar's utterance of the above ascription can now be interpreted as asserting that Odile holds the attitude of belief toward the following explicitly contradictory content:

$$(cv^\wedge CV(<\text{Oscar, 'is witty'}>)) \ \& \ (\neg(cv^\wedge CV(<\text{Oscar, 'is witty'}>)))$$

But Fregean theories are specifically designed to avoid the result that occurrences of ascriptions such as the above attribute to their subjects belief in such "explicit contradictions." In other words, if Fregean theories, in the end, allow for attributions of "explicit contradictions," then it is difficult to see what advantage Fregean theories might have over the Naive Russellian theory. (* I assume that occurrences of 'and' present "cognitive value conjunction," symbolized with '&'. I also assume that

$$cv{\wedge}CV(<Oscar, \text{'is not witty'}>)$$

is equivalent to

$$\neg(cv{\wedge}CV(<Oscar, \text{'is witty'}>))$$

In other words, I also assume that occurrences of 'not' present "cognitive value negation," symbolized with '¬'.)

24 One might attempt to defend a version of the speaker oriented proposal by appealing to similarity relation between contents. That is, one might say that Oscar's utterance of

(3) Odile believes that Twain is a great author.

is true just in case the content Oscar assigns to the that-clause is *sufficiently similar* to the content Odile assigns, or would assign, to the that-clause. This strategy, however, holds little promise. First, how is the relevant *similarity* relation to be defined? Second, such a strategy does not succeed in avoiding any one of the three objections offered here: Under any plausible definition of the similarity relation, it will be possible to formulate semantic arguments from opacity, and there will be nomologically possible cases in which the Echo and Dis-Echo Principles are not respected.

25 The *ceteris paribus* clause is needed to rule out cases in which the subject has revised his beliefs--the principle does not hold if A "changes his mind" concerning the content expressed by the occurrence of Σ. Also, I concede that it is not clear what it is for B to utter an ascription "on the basis of" A's utterance of Σ. This issue, however, is not relevant to my concerns here.

26 I assume that occurrences of sentences suffering from reference failure are neither true nor false. I think this view of occurrences containing empty terms is correct, though the *correctness* of this view is not presupposed by the arguments I present against the subject oriented proposal; i.e. my arguments do not presuppose that on the final analysis occurrences of sentences containing empty terms will turn out to neither true nor false. Some of my arguments against the subject oriented proposal are of the following form: "The subject oriented proposal predicts an instance of reference failure for a particular occurrence, and therefore, if this proposal is correct, our intuitions should dictate that this occurrence is neither true nor false. But our intuitions do not dictate this, and

consequently the subject oriented proposal is false." Thus all that is presupposed by my arguments is that our untutored intuitions, regardless of whether or not these intuitions are correct, dictate that sentences containing empty terms are neither true nor false. This assumption is, I think, indubitable.

27 This tension in Frege's notion of sense is nicely expounded in Evans 1982, Chapter 1. It is important to note that if Evans is correct in maintaining that the notion of sense cannot be extended to solve the puzzles associated with empty terms, then, despite what is often claimed, Russellian theories are no worse off with regard to these puzzles than are Fregean theories; any solution that succeeds for Fregean theories will also succeed for Russellian theories, and *vice versa*.

28 Similar difficulties are presented by occurrences of ascriptions such as

Oscar and Odile believe the Twain is a great author.

Suppose that Oscar and Odile do not "think of Twain in the same way." Whose "way of thinking" is referred to by the occurrence of 'Twain' in the ascription?

29 What is being asked here is, "What *secondary sense* does Oscar employ in thinking of a particular way in which Odile thinks of Twain?" Note that if we are to know what content Oscar's utterance of (3) presents, we must grasp a particular "way" in which Oscar thinks of one of Odile's ways of thinking of Twain. Hence even if Oscar can manage to refer to a particular "way in which Odile thinks of Twain," it seems extremely unlikely that anyone will grasp the thought that Oscar expressed. These difficulties are of course even more pronounced in cases involving iterated attitude ascriptions.

30 The fundamental idea underlying descriptive theories of reference is that, according to such theories, the referent of an occurrence of a referring term is determined by *intrinsic properties* of the speaker; i.e. God, or a clever scientist, could determine the referent of an occurrence by examining the speaker's brain. This idea seems to underlie at least one of Dummett's accounts of sense:

> . . . any one person, if he is to be said to understand ['the Thames'], must be IN COMMAND OF some correct means of identifying the river: if he knows only that 'the Thames' is used as the name of a river, and cannot in any way TELL which river it is the name of, . . . he has only a partial understanding of its sense. (Capitalized words indicate my emphasis. The passage appears in Dummett 1978, p. 99.)

31 Causal theories of reference are bases upon Kripke's causal "picture" of the referent of an occurrence is determined. Kripke summarizes his "picture" of reference as follows:

> An initial "baptism" takes place. Here the object may be named by
> ostension, or the reference of the name may be fixed by a description.
> When the name is "passed from link to link," the receiver of the name
> must, I think, intend when he learns it to use it with the same reference as
> the man from whom he heard it. If I hear the name 'Napoleon' and decide
> it would be a nice name for my pet aardvark, I do not satisfy this
> condition, (Kripke 1972, p. 96).

(Note, first, that Kripke does not distinguish "type reference" from "occurrence reference," and second, as Kripke points out, the above "picture" cannot serve as an eliminative definition, as it appeals to the notion of *intending* to use a term to refer to the same referent.)

32 For a sketch of such a "mixed theory of reference," see Evans 1977.

33 One might object that if Odile believes that CV1 and CV2 determine the same referent, then CV1 and CV2 must be identical. there are several reasons for rejecting such a proposal. First, under this proposal Odile, who realizes that CV1 and CV2 determined the same referent, could not grasp the content presented by Otho's utterance of (1), where Otho does not realize that CV1 and CV2 determine the same referent. Second, it might be that, despite what Odile believes, CV1 and CV2 determine distinct referents, in which case they could hardly be identical. And finally, the proposal is *ad hoc*, and hardly in keeping with the spirit of the Fregean strategy: surely the Fregean theorist does not want to claim that 'Hesperus' and 'Phosphorus' express the same sense for modern day astronomers.

34 Moreover, as was previously pointed out, Odile may not have had any contact with Twain whatsoever. Under this assumption, there is no appropriate "way of thinking" which might satisfy whatever description Oscar employs, and thus the subject oriented proposal predicts that Oscar's utterance lacks a truth value. this prediction, however, does not accord with our intuitions: our intuitions dictate that if Odile has had no contact with Twain whatsoever, if she has "never even heard of him," then Oscar's utterance of (3) is *false*.

35 Thus the arguments advanced in this chapter are in a way analogous to the arguments against Fregean theories advanced by Kripke, Putnam, and Kaplan: One of the central purposes of the arguments advanced by these Philosophers is to show that senses, or cognitive values, cannot both determine referents, or semantic values, and also serve as constituents of the contents expressed by occurrences of declarative sentences. Analogously, the arguments presented in this chapter attempt to show that senses, or cognitive values, cannot both serve to preclude the epistemological and semantic arguments from opacity, and also serve as (i) constituents of publicly accessible contents presented by occurrences of declarative sentences, and (ii) the referents of terms appearing in the that-clauses of attitude ascriptions.

3.O Criticism of Three Russellian Theories.

In this chapter I will explicate and criticize the Russellian theories proposed by Salmon, Richard, and Crimmins and Perry.[1] Though these theories are Russellian in that they can be interpreted as respecting the principles of semantic innocence and direct reference, each of them rejects at least one of the tenets of The Naive Russellian Theory. Recall that The Naive Russellian Theory maintains, in addition to the principles of Direct Reference and Semantic Innocence, the following three general tenets. First,

> The Principle of Full Articulation: The proposition presented by an occurrence of a declarative sentence is wholly determined by (a) the semantic values--the referents and designations--of the phonetically or orthographically realized terms occurring in the sentence, (b) the logical form of the sentence.

And second,

> The Principle of Propositional Truth: The truth conditions of an occurrence of a declarative sentence are identical to, and determined by, the truth conditions of the proposition presented by the occurrence; i.e. no two occurrences which present the same proposition can have distinct truth conditions.

And finally third, The Naive Russellian Theory also endorses a binary analysis of propositional attitudes and propositional attitude verbs: propositional attitudes--the actual mental phenomena--are binary relations between agents and Russellian propositions and, furthermore, occurrences of propositional attitude verbs designate these binary relations. Further recall that the phenomenon of opacity was shown to pose three central difficulties for The Naive Russellian Theory: epistemological arguments from opacity, the problem of accounting for the cognitive significance of

occurrences, and the semantic arguments from opacity. It is because of these difficulties that each of the Russellian theories considered in this chapter rejects at least one of the above tenets of The Naive Russellian Theory.

This chapter proceeds as follows: For each of the three Russellian Theories there are two sections, an expository section and a critical section. The expository sections will focus on how the theories endeavor to avoid the problems posed by opacity; i.e. in the expository sections I will attempt to answer the question, "How does the theory endeavor to preclude the epistemological and semantic arguments from opacity, and also account for the cognitive significance of occurrences?" In the critical sections I will raise objections to the theories, most of which are variations on a similar theme: Each of the theories examined is a Russellian theory; i.e. each theory represents an attempt to preserve the principles of Direct Reference and Semantic Innocence, while avoiding the arguments and problems posed by the phenomenon of opacity.[2] Consequently, none of these theories utilizes the Fregean strategy for avoiding the problems posed by the phenomenon of opacity. Nonetheless, the strategies employed by these Russellian theories are similar enough to the Fregean strategy to be caught between the Scylla and Charybdis of the Fregean strategy: To avoid the epistemological arguments from opacity and account for the cognitive significance of occurrences, each of the Russellian theories posits a kind of entity, a *mediator*, which serves to individuate the relata of propositional attitudes; each of the theories analyzes propositional attitudes--the mental phenomena--as *ternary* relations between agents, Russellian propositions, and some kind of mediator. But in order to avoid refutation by epistemological arguments from opacity and account for the cognitive significance of occurrences, these mediators must be individuated extremely finely, and they must be individuated independently of the phenomenon of opacity. As a consequence, these Russellian theories are threatened by the Fregean strategy's Scylla. Moreover, these Russellian theorists also assume, like the Fregean theorist, that these finely individuated

mediators must somehow be incorporated into the propositions presented by occurrences of attitude ascriptions if the truth conditions of occurrences of ascriptions such as

(3) Odile believes that Twain is a great author.

and

(4) Odile believes that Clemens is a great author.

are to be distinguished. In making this assumption, however, the Russellian theorists become vulnerable to the second half of the Fregean strategy's Charybdis, viz. the problem of preserving the legitimacy of our ordinary attitude ascribing practices. As was argued in section 2.3.2, the semantics of ordinary attitude ascriptions do not seem to involve appeal to anything like finely individuated *mediators* of propositional attitudes.

3.1.1 Salmon's Russellian Theory.

Salmon's theory differs from The Naive Russellian Theory in that it rejects the binary analysis of propositional attitudes and also, strictly speaking, denies that occurrences of propositional attitude verbs designate propositional attitudes. According to Salmon's theory, propositional attitudes--the mental phenomena--are ternary relations between agents, Russellian propositions, and what Salmon calls "modes of apprehension of Russellian propositions." (Thus "modes of apprehension" are the *mediators* posited by Salmon's theory.) Propositional attitude verbs, however, do not designate these ternary relations; rather occurrences of propositional attitude

verbs designate binary relations which are formed by existential generalizations on instances of ternary propositional attitudes. For example, Salmon maintains that Odile's holding the attitude of belief toward the Russellian proposition that Twain is a great author is to be analyzed as follows:

(G) BEL <Odile, P, MA>

where 'P' designates the Russellian proposition that Twain is a great author, and 'MA' designates the "mode of apprehension" by which Odile "grasps" P. (I continue to ignore issues concerning time and tense.) Thus the attitude of belief is identified with the ternary BEL relation where an agent holds the attitude of belief toward a Russellian proposition p just in case the agent has (something like) a "disposition to inward assent [to p] when taken in such and such a way."[3] According to Salmon's theory of attitude ascriptions, however, 'believes' as it appears in occurrences of

(3) Odile believes that Twain is a great author.

does not designate the ternary BEL relation. Rather Salmon maintains that 'believes' as it appears in occurrences of (3) designates an existential generalization of the ternary BEL relation. More specifically, Salmon would analyze an occurrence of (3) as follows:

(3*) ∃m (BEL<Odile, P, m>)

where the variable 'm' ranges over "modes of apprehension."[4]

I now turn to explicating how Salmon's theory endeavors to avoid the problems posed by the phenomenon of opacity. I will begin by considering the epistemological

arguments from opacity. Since Salmon maintains that propositional attitudes--the mental phenomena--are ternary relations between agents, Russellian propositions, and "modes of apprehension of Russellian propositions," the epistemological principles of assent and dissent must be amended:

> Salmon's Epistemological Principle of Assent: If, at time t, a normal, sincere, understanding subject A assents to an occurrence of a declarative sentence Σ which presents Russellian proposition p, then there is some mode of apprehension MA such that in assenting to the occurrence of Σ, A apprehends p via MA, and BEL<A, p, MA>, at t.

> Salmon's Epistemological Principle of Dissent: If, at time t, a normal, sincere, understanding subject A dissents from an occurrence of a declarative sentence Σ which presents Russellian proposition p, then there is some mode of apprehension MA such that in dissenting from the occurrences of Σ, A apprehends p via MA, and ¬(BEL <A, p, MA>), at time t. (Again, *dissenting from* an occurrence of a sentence is to be sharply distinguished from *assenting to the negation* of an occurrence of a sentence.)

Hence, according to Salmon's theory, what follows from Odile's assent to (O1) is not that

(A) Believes <Odile, P>

,

but rather that

(H) BEL <Odile, P, MA$_1$>

(where 'MA$_1$' designates the "mode of apprehension" by which Odile--in assenting to (O1)--apprehends the Russellian proposition P.) And what follows from Odile's dissent from (O2) and the epistemological principle of dissent is not that

$$(B) \quad \neg(\text{Believes } <\text{Odile, P}>)$$

but rather that

$$(I) \quad \neg(\text{BEL } <\text{Odile, P, MA}_2>)$$

(where 'MA$_2$' designates the "mode of apprehension" by which Odile--in assenting to
(O2)--apprehends the Russellian proposition P. I again ignore temporal factors.) Since
(G) and (H) are not contradictories, Salmon's Russellian theory is seemingly able to
preclude refutation by way of epistemological arguments from opacity. (Note that
Salmon's theory succeeds in precluding this argument from opacity only if MA$_1$ is not
identical to MA$_2$.)

How does Salmon's theory account for the cognitive significance of
occurrences? In answering this question it will be useful to compare Salmon's theory
with Frege's theory of sense and reference. According to Frege's theory, the cognitive
significance of an occurrence, relative to an understanding subject, is determined by the
content expressed by that occurrence. Frege maintains that the reason that Odile
assents to (O1) is that (O1) expresses thought T for Odile, and Odile holds the attitude
of belief toward this content. Thus, according to Frege's theory, the cognitive
significance of an occurrence, relative to a subject, is determined by the content
expressed by the occurrence relative to that subject. Salmon, however, denies that the
cognitive significance of an occurrence is determined by the content expressed by the
occurrence. Rather Salmon maintains that the cognitive significance of an occurrence,
for a subject, is determined by the "mode of apprehension" by which the subject grasps
the Russellian proposition expressed by the occurrence. Thus, according to Salmon's
theory, Odile assents to (O1), yet dissents from (O2) because she grasps the Russellian

112

propositions presented by these occurrences via different "modes of apprehension"; it is not the content toward which Odile holds the attitude of belief which accounts for her verbal behavior, but rather it is how--via what mode of presentation--she BELs a particular content which accounts for her behavior.[5]

Salmon's strategy for avoiding the epistemological arguments from opacity and accounting for the cognitive significance of occurrences is thus crucially dependent on there being appropriately individuated *mediators*, or "modes of apprehension of Russellian propositions," and even Salmon concedes that it is not clear what such entities might be.[6] It is clear, however, that whatever Salmon's "modes of apprehension" are, they play a role very similar to the role played by Fregean senses and thoughts. Consider the following passage in which Salmon explains the notion of a "mode of apprehension of a Russellian proposition" by way of contrasting his theory with Frege's theory of sense and reference:

> On the Fregean conception, every piece of cognitive information, every "thought," is made entirely of things like concepts. . . . To apprehend such a "thought" is, it seems, to be fully acquainted with it. There is no changing appearance, no superficial surface concealing the soul, no guise or veil of outward manifestation interceding between the subject and the thing-in-itself. To apprehend it is, as is true of a singular proposition whose only constituent other than things like concepts is a particular sensation or visual sense datum, an item of "direct acquaintance." There is no "failing to recognize" a particular pain, for example by mistaking it for someone else's tickle. To have such a sensation or sense datum is to be acquainted with it in the fullest and most complete way possible. But the modified naive theory allows for propositions of a different sort: singular propositions involving external individuals and material objects as constituents. Clearly, the mode of apprehension for such propositions must be more complex than the mere grasping of pure concepts and the experiencing of wholly internal sensations. Apprehending such a proposition cannot be a wholly

internal, mental act. The means by which one is acquainted with a singular proposition includes as a part the means by which one is familiar with the individual constituents of that proposition. The mode of acquaintance by which one is familiar with a particular object is part of the mode of apprehension by which one grasps a singular proposition involving that object.[7]

.

This passage illustrates that Salmon's "modes of acquaintance" and "modes of apprehension" are indeed close cousins to Frege's senses and thoughts, respectively. There are three striking similarities: First, modes of apprehension, like Fregean thoughts, are composed of *intermediary objects of acquaintance*: On Frege's theory, one does not think of Twain directly, but only "through" a sense with which one is in some sense "directly acquainted." Since senses are constituents of thoughts, thoughts can be distinguished by appeal to senses; thoughts containing distinct senses are themselves distinct. Similarly on Salmon's theory one can think about Twain only "through" some "mode of acquaintance" with which one is again in some sense directly acquainted. Furthermore, modes of apprehension, which are partially composed of modes of acquaintance, can be distinguished by appeal to the modes of acquaintance they contain.

Second, though on Salmon's theory modes of acquaintance are not constituents of contents, Salmon's method for avoiding the epistemological arguments from opacity is structurally very similar to Frege's method. Recall that Frege's theory precludes the epistemological arguments from opacity by positing senses to serve as the contents expressed by occurrences of terms, thereby individuating the contents expressed by occurrences of sentences--thoughts--more finely than Russellian propositions. Concerning the instance of opacity involving Odile, (O1), and (O2), the Fregean theorist claims that what follows from Odile's assent to (O1) and dissent from (O2) is that

(B) Believes <Odile, T>

(where 'T' designates the Fregean thought expressed by (O1).) And what follows from Odile's dissent from (O2) is that

(C) ¬(Believes <Odile, T*>)

(where 'T*' is the Fregean thought presented by (O2).) Salmon's method of avoiding the epistemological arguments from opacity differs from Frege's method only in that, where Frege has the binary believes relation and thoughts, Salmon has the ternary BEL relation and "Russellian propositions under modes of apprehension." Where Frege distinguishes T and T* by appeal to distinct senses of Twain, Salmon distinguishes MA_1 and MA_2 by appeal to distinct "modes of acquaintance of Twain." The structural similarity can be further elucidated if we identify "Russellian propositions under modes of apprehension" as ordered pairs of Russellian propositions and modes of apprehension, and reinterpret the BEL relation as a binary relation between agents and such ordered pairs. What is the theoretical difference between a Fregean thought, and an ordered pair consisting of a Russellian proposition and a "mode of apprehension" of that proposition?˙ What reason could there be for preferring Salmon's response to the epistemological arguments over Frege's response? (If, as was suggested in section 2.3.2, the cognitive value expressed by an occurrence of a term cannot be specified independently of the referent of the term, then it is especially difficult to see how Salmon's method of avoiding the epistemological arguments from opacity differs from the Fregean theorist's method.)

Finally, the third striking similarity between Frege's senses and thoughts, and Salmon's modes of acquaintance and modes of apprehension, is that both kinds of entities are *multiply realizable*: where senses and thoughts are abstract objects which

can be "grasped" by different subjects at different times, modes of acquaintance and modes of apprehension are universals which can be *instantiated* by different subjects at different times.[8] Because both Fregean senses and thoughts, and Salmon's modes of acquaintance and modes of apprehension are *multiply realizable* in this way, both kinds of entities can support nomological generalizations, and thus both kinds of entities can be appealed to in order to account for the cognitive significance of occurrences. For example, according to Frege's theory, Odile assents to (O1) because it expresses a particular thought, T, for her. Implicit in this explanation of Odile's behavior is an appeal to a general psychological law of the form

> For all subjects X, if conditions C obtain, and X perceives a token of a
> sentence of Σ which expresses thought T for X, then *ceteris paribus* X
> will assent.

(The conditions C will presumably specify, among many other things, that X holds the attitude of belief toward T.) There can be such a law only if T is *multiply realizable* in the relevant sense; if the fact that (O1) expresses T for Odile is to account for Odile's assent to (O1), then T must be such that different subjects could grasp T, and Odile might not have grasped T. Similar remarks apply to Salmon's theory. If modes of apprehension are to account for the cognitive significance of occurrences, they must enter into such psychological laws, and consequently they must be *multiply realizable*-- they must be capable of being *intantiated* by different subjects at different times.

It is clear that, with regard to the epistemological arguments from opacity and the problem of accounting for the cognitive significance of occurrences, Salmon's theory is very similar to Fregean theories; with regard to the epistemological arguments from opacity and the problem of cognitive significance the mediators posited by Salmon's theory play roles very similar to the roles played by Fregean senses and thoughts. With regard to the semantic arguments from opacity, however, Salmon's

theory diverges drastically from the Fregean strategy. (That Salmon's theory diverges from Fregean theories is to be expected: it would be surprising if Salmon's theory could both satisfy the Russellian desiderata, and also mimic the Fregean strategy with regard to all three of the problems posed by the phenomenon of opacity.) Recall the semantic principles of assent and dissent:

> The Semantic Principle of Assent: If a normal, sincere, understanding subject assents to an occurrence of a declarative sentence Σ in a context c, then an occurrence of ⌜N believes that Σ⌝, where N refers to the subject, is true in any context c', where c' is semantically similar to c with regard to Σ and ⌜N believes that Σ⌝ .

> The Semantic Principle of Dissent: If a normal, sincere, understanding subject dissents from an occurrence of a declarative sentence Σ in a context c, then an occurrence of ⌜N does not believe that Σ⌝, where N refers to the subject, is true in any context c' where c' is semantically similar to c with regard to Σ and ⌜N does not believe that Σ⌝.

It follows from Odile's assent to (O1) and the Semantic Principle of Assent that an occurrence of

> (3) Odile believes that Twain is a great author.

(in a semantically similar context) is true. Salmon analyzes occurrences of (3) and also occurrences of

> (4) Odile believes that Clemens is a great author.

as follows:

> (3&4*) ∃m (BEL<Odile, P, m>)

Thus according to Salmon's analysis of ascriptions, just as with The Naive Russellian Theory, an occurrence of (3) is true in a context c if and only if an occurrence of (4) is true in c. It follows from Odile's dissent from (O2) and the Semantic Principle of Dissent, however, that an occurrence of

(4n) Odile does not believe that Clemens is a great author.

is true in c (where c is semantically similar to the context of (O1) and (O2). But (4n) is simply the negation of (4), and consequently they cannot both be true relative to the same context of utterance, c. Salmon's theory can thus be refuted by semantic arguments from opacity.

The source of the difficulty is that, under Salmon's analysis of attitude ascriptions, the *mediators* posited in order to preclude the epistemological arguments from opacity are not in any way *specified* by the corresponding attitude ascriptions. The situation involving Odile and (O1) and (O2) does not serve an epistemological argument from opacity against Salmon's theory because of MA_1 and MA_2; so long as MA_1 and MA_2 are distinct, (H) and (I) are not contradictories. According to Salmon's analysis of (3) and (4), however, MA_1 and MA_2 are not incorporated into the propositions presented by occurrences of (3) and (4). Indeed, if Salmon desires to maintain the Principle of Semantic Innocence and the Principle of Full Articulation, how could he maintain that MA_1 and MA_2 are somehow *specified* by occurrences of (3) and (4), respectively? There are no phonetically or orthographically realized terms appearing in occurrences of (3) and (4) that could have MA_1 and MA_2, respectively, as their semantic values, and consequently Salmon is compelled to maintain that occurrences of (3) and (4) present the same proposition. As a result, because he also endorses the Principle of Propositional Truth, Salmon is compelled to maintain that

occurrences of (3) and (4) have the same truth conditions. In summary, if Salmon maintains the Russellian desiderata, as well as the principles of full articulation and propositional truth, he is compelled to maintain that occurrences of (3) and (4) have the same truth conditions, and consequently Salmon's theory is susceptible to refutation by the semantic arguments from opacity.

Salmon is well aware that his theory is unable to preclude the semantic arguments from opacity. His response to these arguments is to deny the Semantic Principle of Dissent. Salmon denies, for example, that it follows from Odile's dissent from (O2) that an occurrence of

(4n) Odile does not believe that Clemens is a great author.

(in a semantically similar context) is true. Recall that Salmon analyzes an occurrence of (4n) as

(4n*) $\neg(\exists m \, (BEL<Odile, P, m>))$

(where the variable 'm' ranges over "modes of apprehension"). But, relative to our situation involving Odile, (O1) and (O2), an occurrence of (4n*) would clearly be false: an occurrence of (4n*) is true only if there is no "mode of apprehension" such that Odile "inwardly assents to P under that mode apprehension." But if Odile assents to (O1), then there is *some* "mode of apprehension" such that Odile "inwardly assents" to P under that mode apprehension, viz. the mode of apprehension displayed, for Odile, by (O1). Thus, according to Salmon's theory, what follows from Odile's assent to (O1) is not that relevant occurrences of (4), or (4n*) are true, but rather that relevant occurrences of

$$(4n^{*'}) \quad \neg(\text{BEL} <\text{Odile}, P, MA_2>)$$

are true. But there is no non circuitous way to express what is expressed by occurrences of (4n*') in ordinary English; the closest we can come is something like, "Odile does not believe that Twain is a great author in the way that proposition is presented by 'Clemens is a great author'."

The Semantic Principle of Dissent, however, is supported by our untutored intuitions concerning the truth conditions of attitude ascriptions, and thus Salmon must provide some explanation as to why our untutored intuitions go astray. Salmon suggests three kinds of reasons in support of his claim that our untutored intuitions concerning the truth conditions of ascriptions such as (4) and (4n) are incorrect.[9]

First, ordinary speakers have "a tendency to confuse" ascriptions such as (4) and (4n) with metalinguistic ascriptions such as

(4ml) Odile believes that occurrences of 'Clemens is a great author' are true.

Given that Odile sincerely dissents from (O2), occurrences of (4ml) of course are false. Thus Salmon maintains that at least part of the reason that our untutored intuitions go astray is that we confuse what is presented by occurrences of (4) with what is presented by occurrences of (4ml); occurrences of (4ml) are false, and thus our tendency to confuse (4) with (4ml) leads us to think that occurrences of (4) are also false. (Similar remarks apply to occurrences of (4n), and the negation of (4ml); we confuse occurrences of (4n), which are false, with occurrences of the negation of (4ml), which are true.)

Second, Salmon claims that ordinary speakers (and sophisticated speakers) are apt to incorrectly infer from Odile's dissent from (O2) that occurrences of

(4n) Odile does not believe that Clemens is a great author.

are true, because ordinary (and sophisticated) speakers over generalize. Salmon

maintains that Odile's dissent from (O2) implies that

$$(J) \quad \exists m \, (\neg(BEL<Odile, P, m>)).$$

There is some "mode of apprehension" by which Odile grasps P, such that Odile is not

"disposed to inward assent to P" when she grasps P under this mode of apprehension,

viz. the mode of apprehension by which Odile grasps P in dissenting from (O2). Thus

Odile's dissent from (O2) implies that (J), (though there is no straightforward way to

express (J) in English). According to Salmon's analysis of attitude ascriptions,

however, (J) does not imply that relevant occurrences of

are true. Recall that occurrences of (4n) are analyzed as having the same truth

conditions as occurrences of

$$(4n^*) \quad \neg(\exists m \, (BEL<Odile, P, m>)).$$

(J) obviously does not imply that relevant occurrences of (4n*) are true: Odile's dissent

from (O2) does imply that there is some mode of apprehending P such that she does not

BEL that P under that mode of apprehension. Her dissent does not, however, imply

that *there is no* mode of apprehending P such that she BELs that P under that mode of

apprehension. (Note that (4n*) is obtained from (J) by replacing the existential

quantifier with a universal quantifier.) Therefore to assume that the truth of

occurrences of (4n) follows from Odile's dissent from (O2)--or more generally, to

endorse the Semantic Principle of Dissent--is to commit the error of over

generalization. Our intuitions concerning the truth conditions of occurrences of (4n) go

astray because we mistakenly assume that if follows from Odile's dissent from (O2)

that Odile does not BEL that P under any mode of apprehension whatsoever, while all

that actually follows from Odile dissent is that there is one mode of apprehending P

such that Odile does not BEL that P via that mode of apprehension. (Note that this

second point concerning over generalization helps to "explain away" our intuition that

occurrences of (4n) are true, though it does not help to "explain away" our intuition that

occurrences of (4) are false.)

Salmon's third, and I think most compelling, reason for supposing that our

intuitions concerning the truth conditions of occurrences of (4) and (4n) go astray

concerns the distinction between semantically encoded information and pragmatically

imparted information. Salmon claims that the reason that ordinary speakers are prone

to incorrectly judge that occurrences of (4) are false (and that occurrences of (4n) are

true) is that they confuse these two kinds of information. Salmon explains,

> where someone under discussion has conflicting attitudes toward a
> single proposition that he or she takes to be two independent
> propositions, . . . there is an established practice of using belief
> attributions to convey not only the proposition agreed to (which is
> specified by the belief attribution) but also the way the subject of the
> attribution takes the proposition in agreeing to it (which is no part of the
> semantic content of the belief attribution).[10]

Thus the reason that ordinary speakers incorrectly intuit that occurrences of (4) are

false is that there is an established practice of using (4) to *pragmatically convey* more

than it presents, or *semantically encodes*: An occurrence of (4) presents, or encodes,

the Russellian proposition that there is some mode of apprehension or other such that

Odile BELs the Russellian proposition P under that mode of apprehension. But, according to Salmon, there is an established practice of using (4) to *pragmatically convey* information concerning which particular mode of apprehension Odile employs in holding the BEL relation toward P. In the following passage, Salmon uses the example of the ancient astronomer to illustrate the distinction between pragmatically conveyed and semantically encoded information:

> The ancient astronomer agrees to the proposition about the planet Venus that it is it when he takes it in the way it is presented to him through the logically valid sentence 'Hesperus is Hesperus' but he does not agree to this same proposition when he takes it in the way it is presented to him through the logically contingent sentence 'Hesperus is Phosphorus'. The fact that he agrees to it at all is, strictly speaking, sufficient for both the sentence 'The astronomer believes that Hesperus is Hesperus' and the sentence 'The astronomer believes that Hesperus is Phosphorus'. Though the sentences are materially equivalent, and even modally equivalent . . . there is a sense in which the first is better than the second, given our normal purpose in attributing belief. Both sentences state the same fact (that the astronomer agrees to the singular proposition in question), but the first sentence also manages to convey *how* the astronomer agrees to the proposition. Indeed, the second sentence, though true, is in some sense inappropriate; it is positively misleading in the way it (correctly) specifies the content of the astronomer's belief. It specifies the content by means of a 'that'-clause that presents the proposition in the "wrong way," a way of taking the proposition with respect to which the astronomer does not assent to it. *This does not affect the truth value of the second sentence, for it is no part of semantic content of the sentence to specify the way the astronomer takes the proposition when he agrees to it* [emphasis added]. The 'that'-clause is there only to specify the proposition believed. It happens in the 'Hesperus'-'Phosphorus' type of case that the clause used to specify the believed proposition also carries with it a particular way in which the believer takes the proposition, a particular x by means of which he or she is familiar with the proposition. In these cases, the guise or appearance by means of which the believer would be familiar with a proposition at a particular time t were it presented to him or her through a particular sentence is a function of the believer and the sentence. Let us call this function F_t. For example, $F_t(x, S)$ might be the way x would take the information content of S, at t, were it presented to him or her through the very sentence S. In the case of the ancient astronomer, we have

BEL<the astronomer, that Hesperus is Hesperus,

F_t(the astronomer, 'Hesperus is Hesperus')>

. . . but not . . .

BEL<the astronomer, that Hesperus is Hesperus,

F_t(the astronomer, 'Hesperus is Phosphorus')>.[11]

Thus Salmon maintains that our untutored intuitions incorrectly dictate that occurrences of (4) are false (and that occurrences of (4n) are true) because we systematically confuse pragmatically imparted information with semantically encoded information: the information pragmatically conveyed by an occurrence is not part of the proposition presented by the occurrence and therefore, by the Principle of Propositional Truth, this information is irrelevant to the truth conditions of the occurrence. The information semantically encoded (or presented) by an occurrence, on the other hand, is contained in the Russellian proposition presented by an occurrence, and therefore is relevant to the truth conditions of the occurrence.[12]

3.1.2 Criticism of Salmon's Russellian Theory.

I have two objections to Salmon's theory. My first objection, which is reminiscent of the Fregean strategy's Scylla, concerns the epistemological arguments from opacity, cognitive significance, and the identity conditions for "modes of apprehension." My second objection, reminiscent of the Fregean strategy's Charybdis, concerns the semantic arguments from opacity, and the failure of Salmon's theory to preserve our untutored intuitions concerning attitude ascriptions.

It was shown in section 2.2 that Fregean theories can succeed in precluding the epistemological arguments from opacity and accounting for the cognitive significance of occurrences only if cognitive values are appropriately individuated. It was shown

above, however, that Salmon's method for precluding the epistemological arguments from opacity and accounting for the cognitive significance of occurrences is very similar to Frege's method, and thus it should come as no surprise that Salmon's theory also faces difficulties similar to those discussed in section 2.2. According to Salmon's theory, what follows from Odile's assent to (O1) is that

$$\text{(H) BEL } <\text{Odile, P, MA}_1>$$

(where 'MA$_1$' designates the "mode of apprehension" by which Odile--in assenting to (O1)--apprehends the Russellian proposition P.) And what follows from Odile's dissent from (O2) is that

$$\text{(I) } \neg(\text{BEL } <\text{Odile, P, MA}_2>)$$

(where 'MA$_2$' designates the "mode of apprehension" by which Odile--in dissenting from (O2)--apprehends the Russellian proposition P). Therefore *if* MA$_1$ is not identical to MA$_2$, Salmon's Russellian theory is able to avoid refutation by way of this particular epistemological argument from opacity. However, if Salmon's theory is to avoid refutation by the arguments from opacity and account for the cognitive significance of occurrences, it must be nomologically impossible for there to be an instance of opacity involving a subject A, and occurrences of $\Sigma(\alpha)$ and $\Sigma(\beta)$ such that these occurrences display the same "mode of apprehension" for A. In other words, if Salmon's theory is to be successful in precluding the epistemological arguments from opacity and accounting for the cognitive significance of occurrences, Salmon's theory must satisfy the following version of the Individuation Constraint:

Salmon's Version of the Individuation Constraint: If there is a
nomologically possible instance of opacity involving a subject A, and
occurrences of sentences $\Sigma(\alpha)$ and $\Sigma(\beta)$, then these occurrences must
display distinct "modes of apprehension" for A.

Furthermore, since "modes of apprehension" are also to account for the cognitive

significance of occurrences (E.g. Odile assents to (O1) because it displays MA_1 for her,

and she dissents from (O2) because it displays MA_2 for her.) Salmon's theory must also

satisfy a version of the Independence Constraint:

Salmon's Version of the Independence Constraint: The posited "modes
of apprehension" must be individuated wholly independently of the
phenomenon of opacity; it must be at least in principle possible to state
the individuation conditions for the posited modes of apprehension
without appealing to the phenomenon of opacity.

My first objection to Salmon's theory is that if "modes of apprehension" are as

Salmon seems to conceive of them, then Salmon's theory does not satisfy the above

individuation constraint, and as a result his theory can be refuted by epistemological

arguments from opacity. Salmon never offers a detailed account of "modes of

apprehension" and he concedes that this is a weak point in his theory. It is relatively

clear, however, that Salmon conceives of "modes of apprehension" as event types--

kinds of "graspings"--which are distinguished by appeal to phenomenological

properties. In the examples that Salmon considers, he employs two methods of

individuating "modes of apprehension": first, "modes of apprehension" are

distinguished on the grounds that they contain distinct "modes of acquaintance," and

"modes of acquaintance" are, in turn, distinguished by appeal to phenomenological

properties.[13] Second, "modes of apprehension" are individuated by appeal to sentence

types via the function $F_t(\_,\_)$, from sentences and subjects to modes of apprehension.

(This function was introduced in the passage from Salmon cited above.) I assume that

the value of $F_t(\_,\_)$ for some subject A and sentence Σ is partly determined by the phenomenological properties of A's encounters with tokens of Σ. (I.e. I assume that if the phenomenological properties of A's Σ encounters are sufficiently different from the phenomenological properties of his Σ' encounters, then $F_t(A,\Sigma)$ is distinct from $F_t(A, \Sigma')$.) If my interpretation is correct, if modes of apprehension are event types which are individuated by phenomenological properties, then Salmon can successfully satisfy the Independence Constraint. Phenomenological properties can, I assume, be specified and distinguished without recourse to the phenomenon of opacity. I will now argue, however, that under this interpretation Salmon's theory cannot satisfy the Individuation Constraint.

Let us consider a situation which is based upon Salmon's Superman\Clark Kent example. Suppose that on Monday Lois Lane observes Clark Kent sitting at his desk. At this point she considers an occurrence of

(25) He is fighting a never-ending battle for truth, justice and the American way.

(Suppose that (25) is painted on a sign which hangs above Clark Kent's desk, and that there is an arrow extending from the token of 'He' pointing directly at Clark Kent.) After a moment of careful reflection, Lois determines that she really has no basis for making such a judgment, and she "inwardly dissents" from the occurrence of (25). According to Salmon's version of the Epistemological Principle of Dissent, it follows that

(K) $\neg$BEL <Lois, C, MA_c>

(Where 'C' designates the proposition presented by the occurrence of (25), and 'MA_c' designates the "mode of apprehension" by which Lois "grasps" C.) The following day,

Tuesday, Lois is following Superman who is performing heroic deeds as usual except that, due to complaints from the phone company, he is wearing his office clothes. Not aware that Lois is tracking him, Superman decides to stop in at the office to make a few calls. Lois follows him and is amazed to see the superhero at Clark Kent's desk. While she observes him she again sees the sign, and she again considers an occurrence of (25). After a moment of careful reflection, Lois determines that what the sign says is true, and she "inwardly assents" to the occurrence of (25). According to Salmon's version of the Epistemological Principle of Assent, it follows that

$$(L) \quad BEL <Lois, C, MA_s>.$$

If MA_c is identical to MA_s (if they are identical types), then Salmon's theory is refuted by an epistemological argument from opacity. Is MA_c identical to MA_s? As I interpret Salmon, MA_c can be distinct from MA_s only if there is some phenomenological property PP such that MA_c has PP, but MA_s does not have PP. But what could this phenomenological property be? The event token occurring on Monday seems to be a phenomenological duplicate of the even token occurring on Tuesday, and thus, if event types are individuated by the phenomenological properties of event tokens, it would seem that MA_c must be identical to MA_s. It could even be stipulated that for some span of time the two event tokens are molecule for molecule and phenomenological matches; in both event tokens Clark Kent is sitting in exactly the same position, Lois Lane's visual impression is exactly similar, etc., so that from the moment Lois Lane peeks into Clark Kent's office and begins to consider the occurrence of (25) until the moment she makes her judgment considering what is expressed by this occurrence, the two event tokens are molecule for molecule and phenomenological matches. How can MA_c and MA_s be distinguished in a non arbitrary, non question-begging way?

I conclude that if "modes of apprehension" are event types individuated by appeal to phenomenological properties, then Salmon's theory cannot satisfy the Individuation Constraint, and can therefore be refuted by epistemological arguments from opacity.[14] Since "modes of apprehension"--like Fregean thoughts--are also invoked to account for the cognitive significance of occurrences, the above example also serves to illustrate that Salmon's theory fails to provide an acceptable account of the cognitive significance of occurrences: According to Salmon's theory, on Monday, Lois Lane dissents from the occurrence of (25) because it displays MA_c for her, yet on Tuesday she assents to the occurrence (25) because it displays MA_s for her. If, as I have suggested, MA_c is identical to MA_s, then these explanations are undermined.

My second objection to Salmon's theory concerns his analysis of attitude ascriptions, and the semantic arguments from opacity. Under Salmon's analysis of attitude ascriptions, modes of apprehension are not constituents of the propositions presented by occurrences of ascriptions. As a result Salmon's theory cannot preclude the semantic arguments from opacity. Salmon's response to these arguments is to deny that they are sound by denying the Semantic Principle of Dissent. Denying this principle, however, comes at a price to our untutored intuitions, and I submit that this price is too high. The intuitions in support of the Semantic Principle of Dissent are firmly entrenched: if Odile dissents from (O2), or if she sincerely utters an occurrence of 'I do not believe that Clemens is a great author', then our intuitions dictate that occurrences of

(4) Odile believes that Clemens is a great author.

(in semantically similar contexts) are false, and that occurrences of

(4n) Odile does not believe that Clemens is a great author.

(in semantically similar contexts) are true. These intuitions are so entrenched that they must be taken as data--an adequate account of propositional attitudes and propositional attitude ascriptions must preserve the veracity of these intuitions. Since Salmon's analysis of ascriptions does not preserve these entrenched intuitions, it must be rejected.

In claiming that Salmon's analysis of ascriptions must be rejected, however, I am *not* that claiming that Salmon has failed to explain why, with regard to attitude ascriptions, ordinary speakers intuit the truth conditions they do. Indeed, in this regard I think Salmon's appeal to what he calls "pragmatically imparted information" is roughly correct (or is at least the beginning of a correct explanation). Rather my objection to Salmon's analysis of attitude ascriptions is that Salmon incorrectly assumes that our intuitions are not correct. That is, because Salmon endorses the Principle of Propositional Truth,

> The Principle of Propositional Truth: the proposition presented by (or the "information semantically encoded by") an occurrence of a declarative sentence determines the truth conditions of the occurrence; i.e. no two occurrences which express the same proposition can have distinct truth conditions.

Salmon is compelled to maintain that, since the Russellian proposition presented by occurrences of (4) is true, ordinary speakers must be incorrect in judging occurrences of (4) to be false. My second objection to Salmon's theory is that Salmon assumes, without argument and despite evidence to the contrary, that the Principle of Propositional Truth is true.

Let us describe the phenomenon to be explained as follows: Ordinary speakers who observe Odile's dissent from an occurrence of (O2) (Odile says, "I don't think that's true.") are without exception disposed to judge relevant occurrences of

(4) Odile believes that Clemens is a great author,

to be false.[15] More specifically, if an ordinary speaker observes Odile's sincere dissent

from (O2), he will be disposed to say things like "that's false" when presented with a

token of (4). According to Salmon, however, this occurrence of (4) is true, and in

uttering "that's false" these ordinary speakers speak, and judge, *falsely*. Thus Salmon

readily concedes that his analysis is at odds with the judgments of *truth* and *falsity* that

ordinary speakers make concerning occurrences of attitude ascriptions. But if this is

so, why should it be supposed that Salmon and ordinary speakers are using the terms

'true' and 'false' in the same way? There are two ways of interpreting the apparent

conflict between the untutored intuitions of ordinary speakers and Salmon's analysis of

ascriptions: (i) it is a substantial disagreement, and either Salmon, or ordinary

speakers, are wrong; or (ii) it is not a substantial disagreement because Salmon and

ordinary speakers do not mean the same thing by 'true' and 'false'--they employ distinct

criteria in determining the "truth values" of occurrences of attitude ascriptions. Salmon

assumes that there is a substantial disagreement, and thus he maintains that ordinary

speakers are wrong. But why should we follow Salmon in supposing that there is a

substantial disagreement?

If the Principle of Propositional Truth were true, then there would be a

substantial disagreement between the intuitions of ordinary speakers and Salmon's

theory, and the passages cited in section 3.1.1 make it clear that Salmon endorses the

Principle of Propositional Truth. Salmon, however, gives no arguments in support of

the principle; he gives no arguments in support of the thesis that the Principle of

Propositional Truth encapsulates the criterion ordinary speakers actually employ in

judging the truth values of occurrences. Moreover, given the resistance ordinary

speakers have to altering their judgments concerning the truth values of attitude

ascriptions such as (4) and (4n), it seems that there is at least *prima facie* reason to reject the thesis that ordinary speakers employ the criterion encapsulated in the Principle of Propositional Truth: if one endorses the principles of Direct Reference, Semantic Innocence, and Full Articulation, then one is constrained to maintain that occurrences of (3) and (4) present the same proposition.[16] Nonetheless, ordinary speakers have an entrenched intuition to the effect that occurrences of (3) are true while occurrences of (4) are false. Therefore, given the arguments in support of the principles of Direct Reference and semantic Innocence, and the plausibility of the Principle of Full Articulation, it would seem that the Principle of Propositional Truth is at least open to question.

3.2.1 Richard's Russellian Theory.

Richard's theory of propositional attitudes can be seen as an attempt to improve upon Salmon's theory. Concerning the epistemological arguments from opacity and the cognitive significance of occurrences, Richard's theory is very similar to Salmon's theory: Richard, like Salmon, analyzes propositional attitudes as ternary relations between subjects, Russellian propositions, and a mediating entity. Furthermore, like Salmon, Richard attempts to preclude the epistemological arguments from opacity and account for the cognitive significance of occurrences by appeal to these appropriately individuated mediating entities. Concerning the semantic arguments from opacity, however, Richard's theory differs a great deal from Salmon's theory: Richard's theory is especially designed to preclude the semantic arguments from opacity, and also preserve the legitimacy of our ordinary attitude ascribing practices.

In explicating Richard's theory, which is rather complex, I will first explicate the epistemological, or psychological, component of Richard's theory, and illustrate how it endeavors to preclude the epistemological arguments from opacity and account

for the cognitive significance of occurrences. I will then explicate the semantic component of Richard's theory, and illustrate how it endeavors to preclude the semantic arguments from opacity, while also preserving the legitimacy of our ordinary attitude ascribing practices.

According to Richard's theory, propositional attitudes--the mental phenomena-- are *mediated relations* between subjects and Russellian propositions. Thus, Richard's analysis of propositional attitudes is very similar to Salmon's analysis: where Salmon maintains that propositional attitudes are ternary relations between subjects, Russellian propositions, and "modes of apprehension," Richard maintains that propositional attitudes are ternary relations between subjects, Russellian propositions, and mental representations, or more specifically "sentences of mentalese." For example, Odile's holding the attitude of belief toward the Russellian proposition that Twain is a great author is analyzed as follows:

$$(K^*) \quad \text{BEL}^* <\text{Odile, P, } \sigma>$$

where 'P' designates the Russellian proposition that Twain is a great author, 'σ' designates a sentence of mentalese which presents P, and <Odile, P, σ> satisfies BEL* if and only if subject Odile has a token of mentalese sentence σ appropriately instantiated in her brain.[17] (Again, I ignore temporal factors.) Hence the psychological component presupposes what Richard calls *Psychological Sententialism*, the view that "the psychological states that constitute our attitudes themselves involve some kind of sentential structure."[18] Richard maintains that the psychological states which constitute our having beliefs (desires, etc.) are relations to mental representations; i.e. in order to hold the attitude of belief toward the Russellian proposition p, one must be in the psychological state of having the belief that p, and to be in this psychological state is to have a mental representation--a sentence of

mentalese--which presents p somehow realized in one's brain. (Whether or not a mentalese sentence σ expresses a Russellian proposition p is partly contingent upon certain *external*, contextual, factors; as Putnam and Burge have argued, external facts concerning one's environment and linguistic community are also relevant in determining the content of one's attitudes. From hereafter these external factors will be taken for granted.) Following Richard, we can simplify this assumption by supposing, for the time being, that tokens of mental representations are tokens of natural language sentences written on various "blackboards" in one's head. (This working assumption will later be rescinded. Also, Richard's more refined view is that such things as mental "images" and "current perceptual experiences" also serve as tokens of mental representations.[19] This point will be important later.) Therefore, it will be assumed (for the time being) that to hold the attitude of belief toward the Russellian proposition p is to have written on one's belief blackboard a (neurological) token of a natural language sentence Σ which expresses p.

How does the psychological component of Richard's theory endeavor to preclude the epistemological arguments from opacity? Richard, unfortunately, does not address the epistemological issues directly, but it is relatively clear that his response would be very similar to Salmon's response. Let us consider the instance of opacity involving the ancient Babylonian astronomer and 'Hesperus' and 'Phosphorus'. The ancient astronomer assents to an occurrence of

(26) P is P.

but dissents from an occurrence of

(27) H is P.

where 'H' and 'P' are the terms of Babylonian which translate as 'Hesperus' and 'Phosphorus', respectively. (I assume, for convenience, that the Babylonians used the English 'is'.) As I interpret Richard, he endorses the following versions of the epistemological principles of assent and dissent:

<u>Richard's Version of the Epistemological Principle of Assent</u>: If a sincere reflective understanding speaker assents to an occurrence of Σ at t, then he has a token of Σ written on his belief blackboard, at t.

<u>Richard's Version of the Epistemological Principle of Dissent</u>: If a sincere reflective understanding speaker dissents from an occurrence of Σ at t, then he does not have a token of Σ written on his belief blackboard, at t.[20]

(Note that these principles presuppose that tokens of mentalese just are instances of natural language term types.) It thus follows from the ancient astronomer's assent to the occurrence of (26) and the above version of the Epistemological Principle of Assent that

(K*) BEL* <the ancient astronomer, C, σ>

where 'C' designates the Russellian proposition that Venus is Venus, and 'σ' designates (26), a neurological token of which is written on the ancient astronomer's belief blackboard. And it thus follows from Odile's dissent from (27) and the above version of the Epistemological Principle of Dissent that

(L*) $\neg$(BEL* <the ancient astronomer, C, σ'>)

where 'σ'' designates (27), a neurological token of which is not written on the ancient astronomer's belief blackboard. Since (26) is distinct from (27), (K*) and (L*) are not

contradictories, and thus Richard's theory is able to preclude this epistemological argument from opacity.

How does the epistemological component of Richard's theory account for the cognitive significance of occurrences? This is a difficult question to answer, as Richard never discusses the relationship between a subject's verbal behavior and what sentences of mentalese he has on his belief blackboard. Nonetheless, it is relatively clear that it is the mentalese correlate of an occurrence of a sentence which determines the cognitive significance of the occurrence. In other words, the astronomer assents to the occurrence of (26) *because* he has a neurological token of (26) written on his belief blackboard, and he dissents from the occurrence of (27) *because* he does not have a neurological token of (27) written on his belief blackboard. (A detailed account of the cognitive significance of these occurrences of (26) and (27) for the astronomer would, I assume, involve appeal to sentences such as I (ALMOST) ALWAYS TELL THE TRUTH being written on the astronomer's "desire blackboard," etc.)[21]

The semantic component of Richard's theory is rather complex, and consequently before explicating this component of the theory in detail, I will first give a more general description of it. Richard summarizes the semantic component of his theory as follows:

> Suppose that Maggie thinks that Odile is tired. For Maggie to think that, it is, of course, not necessary that she think to herself, 'Odile is tired'. She could think Germanically to herself, 'Odile ist mude'. Or she might think to herself, 'You are tired', looking at Odile. But for Maggie to think the thought in question, she does need to think of Odile and to think of the property of being tired--and, of course, to think that the one has the other. For Maggie to think that Odile is tired she must have some representation of Odile and of being tired "put together" in an appropriate way. In some sense of 'sentence', she must employ a mental sentence saying that Odile is tired.
>   Suppose I say, 'Maggie thinks that Odile is tired', attempting to tell you what Maggie thinks. My sentence has a part--'that Odile is tired'--that itself has parts representing Odile and being tired. On each

side of the fence--on the side of the speaker and on the side of the thinker--we find the same sort of thing: a sentence, or sentence-like representation, whose various parts pick out objects and properties:
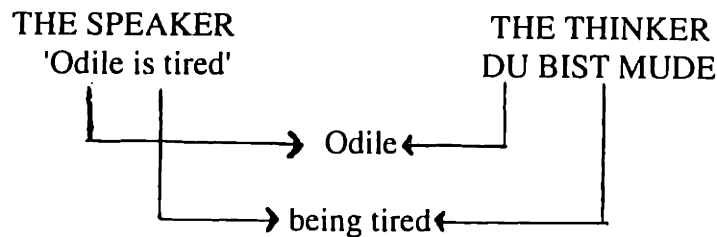
What makes my report true is that my sentence faithfully *represents* a sentence of Maggie's; what makes it false is not faithfully *representing* such, (emphasis added).[22]

It is by invoking the *represents* relation and thereby loosening the constraints on the truth conditions of attitude ascriptions, that Richard hopes to preserve the legitimacy of our ordinary attitude ascribing practices. Theories utilizing the Fregean strategy maintain that the ascription 'Maggie thinks that Odile is tired' is true only if the that-clause of this ascription presents, or *specifies*, something like a Fregean thought, one constituent of which is a cognitive value which Maggie somehow employs in thinking of Odile. The problem with the Fregean theorists' view of ascriptions, according to Richard, is that "they assume that [true] attitude ascription involves a match of non referential cognitive content [i.e. cognitive value] between what the t-clause names and some object of attitude of the person under discussion."[23] On Richard's view there need not be such a *match* between the semantic value of the that-clause, and the content believed (doubted etc.); rather according to Richard's theory it is sufficient for the truth of an occurrence of 'Maggie thinks that Odile is tired' that the that-clause of the ascription in some way *represent* Maggie's mental sentence, DU BIST MUDE.

Furthermore, on Richard's view what mental sentence a that-clause can *represent* is in part determined by the context in which the ascription occurs. Richard explains:

What counts as faithful representation varies from context to context
with our interests and expectations. Context places certain restrictions
on what can represent what. Sometimes these are very specific -- for
example, 'Clark Kent' is to represent only [CLARK KENT]. Sometimes
they are less specific -- a context may say that 'Greg' in 'Maggie wished
that Greg would leave her alone' is to represent terms that connect with
Maggie's current perceptual experience of Greg. Sometimes context is
silent on these matters, and expressions are free to represent any
[mental] expressions with which they corefer. The upshot is that
'Maggie thinks that Odile is tired' is true in a particular context provided
that its [that-clause] represents, according to the context, one of the
sentences which constitutes Maggie's thoughts.[24]

Thus not only is it not necessary that the that-clause of an ascription somehow express
or specify the non referential content grasped by the subject of the ascription, but
whether a that-clause *represents* a subject's mental sentence is contingent upon certain
features of the context of utterance.

I now turn to explicating the semantic component of Richard's theory in more
detail. Richard conceives of the semantic value of a that-clause as being composed of
both the terms (types) contained in the that-clause, and the Russellian proposition
expressed by the occurrence of the that-clause. Richard calls these compound entities
"RAMs" ("Russellian Annotated Matrices"). A RAM is a structure of ordered pairs,
one member of each pair being an expression (type), and the other member being the
Russellian semantic value of that expression. (Richard calls such ordered pairs
"annotations".) Richard maintains that both occurrences of that-clauses and the
sentences of mentalese written on the appropriate blackboards in one's head present (or
determine) RAMs. For example, suppose that Odile has the sentence TWAIN IS A
GREAT AUTHOR written on her belief blackboard. This sentence determines the
following RAM:

Odile's RAM:

        <<IS A GREAT AUTHOR, being a great author> <TWAIN, Twain>>.

Now suppose that Oscar utters

        (3) Odile believes that Twain is a great author.

According to Richard's theory the that-clause of Oscar's utterance determines

Oscar's RAM:

        <<'is a great author', being a great author> <'Twain', Twain>>

For Oscar's utterance of (3) to be true, Oscar's RAM must represent Odile's RAM; or more precisely, for Oscar's utterance of (3) to be true, Oscar's RAM must *represent* a RAM in Odile's "RS" (*representational system*), where Odile's RS contains all the RAMs which are determined by all the sentences of mentalese written on Odile's belief blackboard.

I have stressed that the focal point of Richard's theory is the *represents* relation, and up to this point I have said almost nothing about this relation. What is it for one RAM to *represent* another RAM? (Note that there has been a subtle shift in what represents what; strictly speaking that-clauses do not represent mental sentences; rather, RAMs determined by that-clauses represent RAMs determined by mental sentences.) The only *general* constraint on the represents relation is that if RAM x is to *represent* RAM y, the sentences determining x and y must present the same Russellian proposition. (In other words, x can *represent* y only if, stripped of the linguistic, or

representational, elements of their annotations, x and y are identical Russellian propositions.) Richard gives an account of the *represents* relation, and this general constraint on the relation, in terms of *correlation functions*. A *correlation function* is a function that maps annotations to annotations while preserving reference and designation. For example, a correlation function might map <'Twain', Twain> to <'Sam', Twain>, and <'is a great author', being a great author>, to <'is a terrific writer', being a great writer>, though a correlation function could not map <'Twain', Twain> to <'Dostoyevsky', Dostoyevsky>, nor could a correlation function map <'is a great author, being a great author> to <'is Russian', being Russian>. Therefore, for every correlation function f, if the image of RAM x under f is RAM y, then RAM x and RAM y, stripped of their linguistic elements, would be identical Russellian propositions. A necessary condition for RAM x to represent RAM y can now be stated in terms of correlation functions: RAM x *represents* RAM y only if there is a correlation function f such that $f(x)=y$. In this way the general constraint on the represents relation is enforced.

Though this is the only *general* constraint placed upon the *represents* relation, some occurrences of ascriptions involve other constraints as well. What additional constraints may ! ʌ in operation is determined by the context in which the ascription occurs. This is because, on Richard's view, propositional attitude verbs are *indexicals*. Richard explains:

> In the case of 'that' and 'here', the intentions of the speaker help us
> determine what the uses of the terms refer to. In the case that interests
> us, the intentions of the speaker help determine what a use of 'Twain' in
> 'Odile believes that Twain is dead' can represent. . . . The analogy with
> 'that' and 'here' is simply this: Which possible mediators of Odile's
> belief are relevant to the truth of a use of [(3)] is in part a function of
> various contextually varying factors, including the intentions of the user,
> his interests, and his beliefs about his audience's interests.[25]

Thus according to Richard the semantic value of 'believes' varies from context to context similar to the way in which the semantic values of 'here' and 'that' vary from context to context. Just as 'that' refers to different objects in different contexts, so 'believes' designates different relations in different contexts.[26] This is not to say, however, that 'believes' does not have a *constant meaning*; just as 'here' and 'that' have *characters* which, given a context of utterance, determine a referent, so 'believes' has a character which, given a context, determines a relation. In many contexts we intuit that the that-clause of an ascription need only pick out the appropriate Russellian proposition in order for the ascription to be true. In other contexts, however, meeting this general constraint is not sufficient for the truth of an ascription; in some contexts it is not sufficient for RAM x to *represent* RAM y that x and y, stripped of their linguistic elements, are the same Russellian proposition. In these more *sensitive* contexts *restrictions* are placed upon correlation functions. In other words, 'believes' is an indexical which designates different relations in different contexts, and what changes about the relation designated by 'believes' from context to context is the *restrictions* which are placed on *correlation functions*. Every context c supplies, or induces, a set (perhaps the empty set) of restrictions on correlation functions. A belief ascription occurring in c whose that-clause expresses RAM x is true iff there is a correlation function f which *obeys* the restrictions induced by c and the image of RAM x under f is a RAM in the subject's RS.

What is a *restriction*, and what is it for a correlation function to *obey* a restriction? A *restriction* has three elements: a subject, an annotation, and a set of annotations. For example the following depicts a restriction:

Odile; <'Twain', Twain>; {<TWAIN, Twain>, <CLEMENS, Twain>}

What this says is that in determining the truth value of an ascription concerning Odile, one is restricted to correlation functions which map <'Twain', Twain> to either <TWAIN, Twain> or <CLEMENS, Twain>. For a correlation function f to *obey* this restriction is for f to map <'Twain', Twain> to a member of {<TWAIN, Twain>, <CLEMENS, Twain>}. For example, suppose that the above restriction is induced by the context in which Oscar utters

      (3) Odile believes that Twain is a great author.

Call the RAM determined by the that-clause of (3) 'Oscar's RAM'. Oscar's utterance of (3) is true if and only if there is a correlation function f which obeys the above restriction, (i.e. which maps <'Twain', Twain> to either <TWAIN, Twain> or <CLEMENS, Twain>) and f of Oscar's RAM is a RAM in Odile's RS. More generally, Richard claims that an ascription of the form $\ulcorner$N believes that $\Sigma\urcorner$ is true in a context of utterance c if and only if there is a correlation function f which maps the RAM determined by Σ to a RAM in the RS of the referent of N, and f obeys all of the restrictions which are induced by c.

How does the semantic component of Richard's theory endeavor to preclude the semantic arguments from opacity? Let us again consider an instance of opacity involving the ancient astronomer and 'Hesperus' and 'Phosphorus'. Suppose that the ancient Babylonian astronomer assents to an occurrence of

      (26) P is P.

but dissents from an occurrence of

      (27) H is P.

where 'H' and 'P' are the terms of Babylonian which translate as 'Hesperus' and 'Phosphorus', respectively. If, as we have been assuming, mental terms (tokens) just are terms (types) of natural languages, it follows that the ancient astronomer has a token of (26) written on his belief blackboard, but does not have a token of (27) written on his belief blackboard. Furthermore, according to Richard's theory there being a token of a sentence on a subject's belief blackboard is both necessary and sufficient for that subject's having the RAM determined by that sentence in his RS. Thus the ancient astronomer has

P=P RAM: <<'is', identity>, <<'P', Venus>, <'P', Venus>>>

in his RS but does not have

H=P RAM: <<'is', identity>, <<'H', Venus>, <'P', Venus>>>.

in his RS. Since RAMs are individuated as finely as the expression types occurring in them, the P=P RAM is not identical to the H=P RAM.

Suppose that in a particular context c, our naive intuitions dictate that an occurrence of

(28) The ancient astronomer believes that Phosphorus is Phosphorus.

is true, while our intuitions dictate that, in c, an occurrence of

(29) The ancient astronomer believes that Phosphorus is Hesperus.

143

is false. Richard would have it that, in such a *sensitive* context, the following restriction is induced:

The astronomer; <'Phosphorus', Venus>; {<'P', Venus>}
<'Hesperus', Venus>; {<'H', Venus>}

If context c induces the above restriction, then, since there is a correlation function f which obeys this restriction and also maps the RAM expressed by the that-clause of (28) to a RAM in the astronomer's RS (viz. the P=P RAM), it follows that (28) is true in context c. And since there is no correlation function which obeys the above restriction and maps the RAM presented by the that-clause of (29) to a RAM in the astronomer's RS, (29) is false in context c. Thus Richard's analysis avoids refutation by way of this semantic argument from opacity.

Because Richard construes propositional attitude verbs as indexicals so that in different contexts different *restrictions* are placed on what the RAM determined by that-clause can *represent*, Richard's analysis is, unlike Salmon's analysis, seemingly able to preclude the semantic arguments from opacity, without denying the veracity of our untutored intuitions concerning attitude ascriptions. Moreover, Richard's analysis differs from Frege's analysis in that Richard's analysis requires of a true ascription only that its that-clause determine a RAM which *represents* a RAM in the subject's RS; Richard does not require anything like a *match* in cognitive value. Consequently Richard's analysis seems better suited to preserve the legitimacy of our ordinary attitude ascribing practices.

3.2.2 Criticism of Richard's Russellian Theory.

In this section I will present criticisms of both the epistemological and the semantic components of Richard's theory, and my criticisms will again be reminiscent of the Fregean strategy's Scylla and Charybdis. First, I will argue that the epistemological component of Richard's theory does not successfully preclude the epistemological arguments from opacity and, consequently, it also fails to account for the cognitive significance of occurrences. Second, I will argue that--despite the fact that it is specifically designed to do so--the semantic component of Richard's theory fails to preserve the legitimacy of our ordinary attitude ascribing practices.

Richard's strategy for precluding the epistemological arguments from opacity is structurally very similar to Salmon's strategy: both Salmon and Richard analyze propositional attitudes as ternary relations between subjects, Russellian propositions, and a finely individuated, publicly accessible, *mediator*. Furthermore, they both attempt to preclude the epistemological arguments from opacity and account for the cognitive significance of occurrences by appeal to these mediators; the only difference between the theories is that Salmon identifies mediators with "modes of apprehension" while Richard identifies them with sentences of mentalese. Hence, just as Salmon is committed to versions of the individuation and independence constraints, so Richard is committed to versions of these constraints:

> Richard's Version of the Individuation Constraint: If there is a nomologically possible instance of opacity involving a subject A, and occurrences of sentences $\Sigma(\alpha)$ and $\Sigma(\beta)$, then the mentalese correlates of these occurrences for A must be distinguished.

> Richard's Version of the Independence Constraint: The posited sentences of mentalese must be individuated wholly independently of the phenomenon of opacity; it must be at least in principle possible to state the individuation conditions for the posited sentences of mentalese without appealing to the phenomenon of opacity.

In section 3.1.2 I argued that Salmon's theory cannot satisfy the relevant version of the Individuation Constraint; Salmon does not individuate "modes of apprehension" finely enough. My objection to Richard's theory is not that it fails to satisfy the relevant version of the Individuation Constraint, but rather that it fails to satisfy the relevant version of the Independence Constraint: in providing individuation conditions for the mediating entities posited by his theory, Richard, in the end, appeals to the phenomenon of opacity itself.

According to the epistemological component of Richard's theory, what follows from the ancient astronomer's assent to an occurrence of

(26) P is P.

is that

(K*) BEL* <the ancient astronomer, C, σ>

and what follows from the astronomer's dissent from

(27) H is P.

is that

(L*) ¬(BEL* <the ancient astronomer, C, σ'>).

So long as σ is distinct from σ', (K*) and (L*) are not contradictories and Richard's theory successfully precludes the corresponding argument from opacity. We have been

assuming that tokens of mentalese expressions are instances of natural language types, and under this assumption Richard's theory is very similar to a Fregean theory which individuates cognitive values by appeal to natural language expression types. It was shown in section 2.2, however, that Kripke's "Paderewski" cases can be used to refute such Fregean theories, and thus, given the similarity between Richard's theory and such Fregean theories, it is not surprising that "Paderewski" cases also pose problems for Richard's theory.

Suppose that Peter assents to an occurrence of

(17) Paderewski is a talented musician.

but dissents from a distinct occurrence of this same sentence. It follows from Peter's assent to the occurrence of (17) and

> Richard's Version of the Epistemological Principle of Assent: If a
> sincere reflective understanding speaker assents to an occurrence of Σ
> at t, then he has a token of Σ written on his belief blackboard, at t.

that

(M*) BEL* <Peter, Q, σ>.

where 'Q' designates the Russellian proposition that Paderewski is a talented musician. And it follows from Peter's dissent from the other occurrence of (17) and

> Richard's Version of the Epistemological Principle of Dissent: If a
> sincere reflective understanding speaker dissents from an occurrence of
> Σ at t, then he does not have a token of Σ written on his belief
> blackboard, at t.

that

$$(N^*) \quad \neg(BEL^* <Peter, Q, \sigma'>.$$

The problem is of course that if we continue to assume that mental tokens are instances of natural language types, it follows that both σ and σ' are identical to (17), and thus (M*) and (N*) are contradictories.[27]

The essence of the problem is that under the working assumption that mental tokens are instances of natural language types, Richard's theory does not individuate the mediating entities finely enough: if Richard's theory can individuate triples of agents, Russellian propositions, and sentences of mentalese only as finely as natural language expression types, Richard's theory has no way of individuating Peter's "Paderewski the musician" beliefs, from his "Paderewski the statesman" beliefs. As a result, Kripke's "Paderewski" case serves as a kind of instance of opacity with regard to Richard's theory; under the assumption that the mediating entities are natural language types, Richard's theory does not satisfy the individuation constraint. If Richard is to avoid these difficulties he must abandon the assumption that token mental sentences are tokens of natural language types, and provide an *independent* principle of individuation for the mediating entities which individuates them more finely than natural language types.

Richard attempts to do this by amending his theory in the following way: first an equivalence relation, τ *and* τ' *are of the same representational type for x* is defined over natural language and mental expression tokens. The equivalence classes formed under this relation are now identified with, or mapped one-to-one with, *representational types*, and it is these representational types which now serve as the mediating entities.[28] But how are the equivalence classes of natural language and

mentalese tokens determined? What is it for two tokens to be of the same *representational type for a subject x?*

Richard states that there are two sorts of conditions which are necessary and jointly sufficient for two tokens to determine (or be of) the same *representational type*: "outside" conditions, and "inside" conditions. In the case of proper name tokens τ and τ', the outside condition for τ and τ' being of the same representational type is that τ and τ' must be part of the same "causal chain of transmission."[29] Thus it is not enough that tokens τ and τ' be phonetically, orthographically, or even "electro-chemically" similar in order for them to be of the same representational type; e.g. a token of 'Aristotle' used to refer to the shipping magnate is not of the same representational type as a token of 'Aristotle' used to refer to the Greek philosopher. It is clear, however, that the outside condition alone will not solve Richard's problem concerning Peter and Paderewski, as the two tokens of 'Paderewski' appearing in the relevant occurrences of (17) may very well be members of the same "causal chain of transmission." So if Richard is to avoid the recently rehearsed argument from opacity against his view, he must appeal to the inside condition.

The inside condition is what Richard calls "the recognition condition." Roughly speaking, two tokens τ and τ' satisfy *the recognition condition* for *x* if and only if *x* *recognizes* the referent of τ as being the referent of τ'. Richard explains:

> Usually when we hear someone talking about someone, we know (or think we know) who is being talked about. We hear someone say, 'Reagan was going to bomb Nicaragua', and we simply assume that it is the former president who is being discussed, not his wife, son, or daughter, much less an animal rights philosopher, or erstwhile White House chief of staff.

When this is true--when, intuitively speaking, we recognize the subject of discussion--it seems appropriate to think that we are in some sense "filing" or "grouping" the token 'Reagan' we are "processing" along with certain other tokens on our blackboard ("presidential tokens") and segregating it from certain others. In such a case, the new token of 'Reagan' bears the interior relation for sameness of representational type to the presidential tokens.[30]

Since Peter does not *recognize* one token of 'Paderewski' (the token in the sentence to which he assents) as referring to the same individual as is referred to by the other token (the token in the sentence from which he dissents), he does not "group," or "file" the tokens together. Therefore, the "interior" condition for the individuation of representational types dictates that corresponding to the one lexical type 'Paderewski', Peter has two *representational types*. Assuming that the criteria for two tokens being of the same representational type can be extended to cover terms other than names, Richard can now preclude the epistemological argument based upon Kripke's 'Paderewski' case by appeal to representational types: Richard's versions of the epistemological principles of assent and dissent can be amended as follows:

> Richard's Amended Epistemological Principle of Assent: If a sincere, reflective, understanding speaker assents to an occurrence of Σ at t, then he has a token of the same *representational type* as this token of Σ written on his belief blackboard, at t.

> Richard's Amended Epistemological Principle of Dissent: If a sincere, reflective, understanding speaker dissents from an occurrence of Σ at t, then he does not have a token of the same representational type as this token of Σ written on his belief blackboard, at t.

According to these amended principles, what follows from Peter's assent to the occurrence of (17) is that

$$(M^{*\prime}) \quad BEL^* \ <Peter, \ Q, \ RT_\sigma>$$

150

where 'RT$_\sigma$' designates the *representation type* of the occurrence of (17) to which Peter assents. And what follows from Peter's dissent from the distinct occurrence of (17) is that

$$(\text{N*'}) \quad \neg(\text{BEL*} <\text{Peter, Q, RT}_\sigma>$$

(where 'RT$_{\sigma'}$' designates the representational type of the occurrence from which Peter dissents). Is RT$_\sigma$ identical to RT$_{\sigma'}$? I assume that two token sentences Σ and Σ' are of the same representational type if and only if they are isomorphically composed out of constituent terms which are of the same representational type. Let us call the token of (17) to which Peter assents t_σ, and the token from which he dissents $t_{\sigma'}$. Since the token of 'Paderewski' appearing in t_σ and the token of 'Paderewski' appearing in $t_{\sigma'}$ are of distinct representational types, and the representational type of a sentence token is determined by the representational types of the tokens appearing in the sentence, it follows that t_σ and $t_{\sigma'}$ are of distinct representational types. Therefore, since t_σ and $t_{\sigma'}$ are of distinct representational types, and RT$_\sigma$ and RT$_{\sigma'}$ are the representational types of t_σ and $t_{\sigma'}$, respectively, it follows that RT$_\sigma$ and RT$_{\sigma'}$ are not identical.[31]

The problem with Richard's response to the 'Paderewski' instance of opacity is that it violates the *Independence Constraint*. Recall that representational types are equivalence classes defined under the *same representational type for x* relation, and that there are two necessary and jointly sufficient conditions for a name token τ being of the same representational type as a name token τ', for X: first, τ and τ' must be members of the same historical chain of transmission, and second, τ and τ' must be *recognized* by X as referring to the same object. It is the second condition which violates the Independence Constraint. The problem is that we have no *independent* means of determining whether a subject recognizes two tokens as referring the same thing; the

phenomenon of recognition failure is too closely associated with the phenomenon of opacity. (The relation between the two phenomena is further explored in Appendix A.)

Consider again the instance of opacity involving Odile and (O1) and (O2). It is surely correct that in some intuitive sense Odile does not recognize the relevant token of 'Twain' as referring to the same thing as the relevant token of 'Clemens'. But how do we know this? Though Richard calls the recognition condition an "interior" condition, there is nothing "interior" about its application; we have no method whereby we look inside Odile's brain to determine whether or not she "files" the relevant mental tokens together. (Note that if we did possess such a method, then the "recognition condition" would be stated in neuro-physiological terms, and would not involve the concept of *recognition* at all.) What, then, justifies the intuitive claim that Odile does not *recognize* 'Twain' as it appears in (O1) as referring to the same thing as 'Clemens' as it appears in (O2)? I submit that what justifies this claim is simply Odile's behavior with regard to sentences containing the terms 'Twain' and 'Clemens': we intuit that Odile does not recognize the token of 'Twain' appearing in (O1) as referring to the same thing as the token of 'Clemens' appearing in (O2) *because* Odile assents to (O1) yet dissents from (O2), and because she tends to say things like, "I like Twain, but I'm not sure about Clemens," etc. Thus the problem with Richard's recognition condition is that we have no means of determining whether or not a subject recognizes two tokens as referring to the same thing other than by appeal to (what can be construed as) instances of the phenomenon of opacity.[32] Consequently, in invoking the recognition condition Richard is implicitly appealing to the phenomenon of opacity itself, thereby violating the Independence Constraint.[33]

I now turn to criticizing the semantic component of Richard's theory. Richard attempts to preserve the legitimacy of our ordinary attitude ascribing practices by analyzing propositional attitude verbs as complex indexicals. Richard claims, for example, that the semantic value of an occurrence of 'believes' is a relation which holds

between a subject and RAM x if and only if there is a correlation function f which obeys all the restrictions induced by the context of utterance and there is a RAM in the subject's RS which is the image of RAM x under f; since different contexts of utterance will induce different restrictions, the relation designated by 'believes' varies from context to context. It should be noted that the claim that propositional attitude verbs are indexical in this way is an *empirical* claim concerning how ordinary speakers use and understand occurrences of attitude verbs; it is as an empirical claim concerning what is required of a speaker if he is to know the meaning of attitude verbs. Consequently, if Richard's analysis of attitude verbs is correct, it should accord with ordinary speaker's intuitions concerning the meanings of sentences containing attitude verbs. As I will now argue, however, Richard's analysis does not accord with many such intuitions, and this constitutes compelling evidence against Richard's analysis.

If Richard's analysis of propositional attitude verbs were correct, then there could be no "eternal sentences" containing propositional attitude verbs.[34] (An *eternal sentence* is a sentence whose truth conditions do not vary from context to context.) For example, if Richard's view were correct, then one could not determine what propositions are presented by sentences such as

(30) What one believes is one's own business.

and

(31) Hope is all we have.

unless certain contextual features were specified. On Richard's view, (30) and (31) are similar to

(32) I smoke.

and

(33) He's nice.

But this prediction does not accord with the intuitions of ordinary speakers; (30) and (31) are not "incomplete" in anything like the way in which (32) and (33) are incomplete. In order to determine what proposition is presented by an occurrence of (32) one must know who uttered (32) and in order to determine what proposition is presented by an occurrence of (33) one must know which male is most salient in the context (or something along these lines). But occurrences of (30) and (31), unlike (32) and (33), seem to be true or false independently of any such contextual features. (Of course (30) and (31) are probably context dependent because of the tense of the verbs, or the context dependency of pronouns, but these considerations are clearly irrelevant.[35])

Another problem concerns the noun forms of propositional attitude verbs. What, on Richard's theory, is the relationship between 'believes' and 'belief'? Though Richard never discusses the noun forms of attitude verbs, it is clear that they must be interpreted as nonindexical expressions which designate the appropriate epistemological attitudes, as opposed to one of the many possible designatums; i.e. an occurrence of 'belief' must designate the epistemological attitude of belief, and not one of the many possible semantic values of 'believes'. (Richard himself certainly uses 'belief' in this way.[36]) But this bifurcation between the semantic values of the noun and verb forms of propositional attitude expressions does not accord with how ordinary speakers use these expressions. For example, it seems that for any context c, if an occurrence of

(3') Belief is the attitude Odile holds toward the
proposition, or thought, that Twain is a great author.

is true in c, then an occurrence of

(3) Odile believes that Twain is a great author.

is also true in c. But under Richard's analysis (3') no more entails (3) than 'Mary loves

John' entails 'Mary loves *him*'. There are contexts in which 'him' and 'John' are not

coreferential, and in these contexts 'Mary loves John' and 'Mary loves him' may differ

in truth value. Similarly, under Richard's analysis there are contexts in which 'belief'

and 'believes' do not designate the same relation, and thus there are contexts in which

(3') is true, yet (3) is not true. Suppose Odile holds the attitude of belief toward the

Russellian proposition that Twain is a great author because she has a token of the

sentence of mentalese CLEMENS IS A GREAT AUTHOR written on her belief

blackboard, though she does not have a token of the sentence of mentalese TWAIN IS

A GREAT AUTHOR written on her belief blackboard. Further suppose that the

context relative to which (3') and (3) are uttered induces a restriction according to

which <'Twain', Twain> must be mapped to <TWAIN, Twain>. If the context

determined such a restriction, there would be no acceptable correlation function f such

that the image of the RAM expressed by the that-clause of (3) under f was a RAM in

Oscar's RS. Consequently Richard's theory predicts that in some contexts (3') would be

true, while (3) would be false. This prediction, however, does not accord with how

ordinary speakers use expressions of propositional attitudes.

Another problem for Richard's analysis concerns occurrences of iterated attitude

ascriptions. The problem arises because of the following property of indexicals: the

semantic value of an occurrence of an indexical (or a demonstrative) is always

155

determined relative to the context in which it appears. For example, if Oscar utters an occurrence of

(34) Odile believes that today is Monday,

in context c, then the semantic value of 'today' is determined relative to c; 'today' as it appears in Oscar's utterance of (34) refers to the day on which the utterance takes place. Therefore, according to Richard's analysis, two occurrences of 'believes' ('doubts' etc.) will designate the same relation only if the occurrences happen to take place in contexts which are *semantically similar* with regard to 'believes' ('doubts' etc.). This result, however, does not accord with our untutored intuitions concerning the truth conditions of occurrences of iterated attitude ascriptions.

For example, suppose that Oscar sincerely utters

(3) Odile believes that Twain is a great author.

in a context c. Since 'believes' is an indexical, the semantic value of 'believes' as it appears in Oscar's utterance is determined relative to c. Let us call the semantic value determined by Oscar's utterance of 'believes' in c, 'c-bel'. Now suppose that, on the basis of Oscar's utterance of (3), Otho utters an occurrence of

(35) Oscar believes that Odile believes that Twain is a great author

in a distinct context c'. Since the semantic value of an occurrence of 'believes' is determined relative to its context, both appearances of 'believes' in Otho's utterance of (34) designate the same semantic value. Let us call this semantic value, 'c'-bel'. The difficulty arises because c-bel need not be coextensive with c'-bel; there may be many

restrictions in operation in c which are not in operation in c', or vice versa. Let us suppose that this is the case, and thus c-bel is not identical to, or coextensive with, c'-bel. According to Richard's analysis of attitude ascriptions, Otho's utterance of (35) is true just in case the c'-bel relation holds between Oscar, and the following RAM:

RAM (35):

       <<'believes', c'-bel><<'Odile', Odile>, <'Twain is a great author', P>>>

where 'P' designates the Russellian proposition that Twain is a great author. More specifically, Otho's utterance of (35) is true just in case there is correlation function f such that f obeys all the restrictions induced by c' and f(RAM (35)) is a RAM in Oscar's RS. The problem is that, even if c' induces no restrictions at all, there is no such correlation function: since c-bel is distinct from c'-bel, Oscar (we may suppose) does not have any RAMs in his RS which contain the c'-bel relation as a Russellian constituent. The problem is thus that when Oscar utters (3) in c he presents a Russellian proposition p. But when Otho utters (35) on the basis of Oscar's utterance, the that-clause of Otho's utterance does not present p. The problem arises because, according to Richard's analysis, 'believes' designates distinct relations in c an c'.[37]

The above considerations serve as compelling evidence against Richard's analysis of propositional attitude verbs. Richard is correct in pointing out that propositional attitude ascriptions which attribute a particular belief (desire, etc.) to a particular subject seem to exhibit a kind of context sensitivity; our untutored intuitions dictate that in some contexts all that is required for an ascription to be true is that its that-clause present an appropriate Russellian proposition, while in other contexts, sensitive contexts, our untutored intuitions dictate that the that-clause of an ascription must do more than this. Richard attempts to account for this context dependency of

attitude ascriptions by locating the context dependency within the semantics of propositional attitude verbs. The difficulties discussed above arise because propositional attitude expressions--nouns and verbs--are used to do more than to ascribe particular beliefs (desires etc.) to particular subjects, and these other uses of attitude expressions do not seem to involve the kind of context sensitivity posited by Richard's analysis.[38] This suggests that it is a mistake to locate the context sensitivity of attitude ascriptions within the semantics of attitude verbs. (Of course Richard could claim that propositional attitude expressions are ambiguous, e.g., that sometimes propositional attitude verbs are indexicals, and sometimes not, but this would be an *ad hoc* and counterintuitive move.)

I conclude this section with a more general objection to Richard's analysis of attitude ascriptions: Richard's analysis, like the Fregean analysis, does not preserve the legitimacy of our ordinary attitude ascribing practices. Recall that the Fregean analysis encounters this difficulty because it requires ordinary speakers to use the that-clause of their utterances to designate (or *match*) finely individuated cognitive values. This is problematic because in many cases ordinary speakers simply lack the ability to refer, even "by description," to such esoteric entities. Richard's analysis attempts to avoid this difficulty by denying that it is necessary for the truth of an ascription that the that-clause of the ascription match the content of the subject's propositional attitude. Rather, according to Richard's analysis, an ascription is true just in case the that-clause *represents* an appropriate mental representation, and whether or not the that-clause represents an appropriate mental representation is contingent upon whether or not the speaker designates an appropriate believes-relation. In making this move, however, Richard has merely relocated the problem. Richard's analysis suffers from essentially the same difficulties which plague the Fregean analysis of attitude ascriptions, except that in Richard's case the problems do not concern the semantic values of *that-clauses*, but rather arise with regard the semantic values of occurrences of *attitude verbs*.

According to Richard's analysis, propositional attitude verbs are indexicals whose "constant meanings" are characters, functions from contexts to relations. Simplifying somewhat, the character of 'believes' is a function from contexts c, to sets s of ordered pairs $<x,r>$ of subjects x, and RAMs r. More precisely, Richard defines the character of 'believes' (designated by 'CHAR( )') as follows:

> CHAR(c)=s where $<x, r>$ is a member of s iff there is a correlation function f, and a RAM q such that q is a RAM in x's RS and the image of r under f is q, where f obey R(c).[39]

What is R( )? Richard states that R( ) is a "function which takes a context to the set of restrictions it provides."[40] Thus it is R( ) which accounts for the context sensitivity of 'believes'; 'believes' will express distinct relations in contexts c and c' only if R(c) is distinct from R(c'). Thus, if an ordinary speaker, or listener, is to know what is said by an attitude ascription uttered in a context c, he must know the value of R(c); i.e. he must know what restrictions are in operation in c. However, given that determining what restrictions are in operation in a given context will require ordinary speakers to have detailed knowledge concerning one another's mental representations (or more specifically, one another's *representational types*), is it at all plausible to suppose that ordinary speakers, and listeners, have the ability to come to know the value of R( ) for a given context? (Note that, though Richard makes a few sketchy remarks concerning what factors determine the restriction in operation in a given context,[41] he never considers the question of how ordinary speakers might *come to know* what restriction is in operation in this context.)

Suppose that Oscar utters

(3) Odile believes that Twain is a great author.

in a context c. In order for Oscar, and his listeners, to know what proposition Oscar's utterance has presented, in order for them to know what Oscar has said, they must be able to compute the value of R(c). Suppose that R(c) is the following restriction:

Odile; <'Twain', Twain>; {<RT_1, Twain>, <RT_2, Twain>}

where 'RT_1' and 'RT_2' designate *representational types*; i.e. sets of mental representation tokens, including perhaps "mental images," and "current perceptual experiences." (Recall that the epistemological arguments from opacity, specifically "Paderewski" cases, force Richard to use *representational types* as the linguistic elements of annotations.) I submit, however, that it is not at all plausible to suppose that in order for Oscar and his listeners to know what Oscar has said--in order for them to know the truth conditions of Oscar's utterance--they must know that the above restriction is induced by the context of Oscar's utterance. Indeed, I suggest that it is extremely implausible to suppose that Oscar and his listeners--who certainly know the truth conditions of Oscar's utterance--are even capable of coming to know that the above restriction is in operation. Oscar and his listeners, it may be assumed, know next to nothing about, and believe next to nothing about, Odile's mental representations. Consequently, they simply lack the ability to discern that the above restriction, as opposed to some other restriction which contains other representational types, is in operation in the context of Oscar's utterance. And if Oscar and his listeners cannot discern whether restriction A or restriction B is operation, they cannot determine the semantic value of 'believes' as it appears in Oscar's utterance, and consequently they cannot know the truth conditions of Oscar's utterance. But this is unacceptable; ordinary speaker and ordinary listeners typically do know the truth conditions of ordinary occurrences of attitude ascriptions.[42]

Thus another difficulty with Richard's analysis is that it is incompatible with the

ability of ordinary speakers to understand, to know what is said by, occurrences of

attitude ascriptions. That Richard's analysis suffers from this difficulty is rather ironic.

In criticizing the Fregean analysis of attitude ascription, Richard states,

> there is nothing in our day-to-day practice of attitude ascription that
> could be construed as looking for sameness or similarity of sense, as a
> basis for our ascriptions. We just listen and look: if B utters 'Tully is
> tubby' and we think him sincere, we will say, 'He thinks Tully is Tuby'.
> What does similarity of sense have to do with it?[43]

I of course concur with this objection to the Fregean analysis, but the objection applies,

mutatis mutandis, with equal force to Richard's own analysis: There is nothing in our

day-to-day practice of attitude ascription that could be construed as looking for what

restriction is in operation in a given context. We just listen and look. What do

restrictions on what mental representations a that-clause can *represent* have to do with

it?

3.3.1 Crimmins and Perry's Russellian Theory.

Crimmins and Perry's theory differs from the theories of Salmon and Richard in

two important respects. First, though Crimmins and Perry follow Salmon and Richard

in endorsing a ternary analysis of propositional attitudes, Crimmins and Perry do not

identify the third elements of these relations--the *mediators*--with multiply realizable

universals, but instead identify them with *particular* brain-events, or mental

representation *tokens*. Identifying the mediators of propositional attitudes with such

particulars, rather than multiply realizable universals, has important consequences for

Crimmins and Perry's response to the epistemological arguments from opacity, and

their approach to the problem of cognitive significance. Second, Crimmins and Perry's

theory differs from the theories of Salmon and Richard in that Crimmins and Perry

deny that the Principle of Full Articulation holds for occurrences of attitude ascriptions;

i.e. Crimmins and Perry deny that the proposition presented by an occurrence of an

attitude ascription is fully determined by (a) the semantic values--the referents and

designations--of the phonetically or orthographically realized terms appearing in the

sentence, and (b) the logical form of the sentence.[44] Denying the Principle of Full

Articulation is the key feature of Crimmins and Perry's response to the semantic

arguments from opacity. In this section I will first explicate Crimmins and Perry's

response to the epistemological arguments from opacity and their approach to the

problem of cognitive significance. I will then explicate their response to the semantic

arguments from opacity.

The token mental representations which serve as the mediators of propositional

attitudes in Crimmins and Perry's theory are called *cognitive particulars*. Individual

instances of propositional attitudes--particular *beliefs, desires*, etc.--are kinds of

cognitive particulars. Thus, since cognitive particulars are *tokens*, as Crimmins and

Perry use the term it is impossible for two individuals to have (or instantiate) the same

belief. Cognitive particulars also have structure and content. They are structured

entities composed of *notions* and *ideas*, where notions are mental representations

(tokens) of objects, and ideas are mental representations (tokens) of n-ary relations.

The *content* of a notion is an object and the *content* of an idea is an n-ary relation.

Thus the content of a belief is the Russellian proposition determined by (i) the content

of the notions and ideas which make up the belief, and (ii) the structure of the belief.[45]

Hence, according to Crimmins and Perry's analysis of propositional attitudes, for Odile

to hold the attitude of belief toward the Russellian proposition that Twain is a great

author is for Odile to instantiate a belief which is appropriately composed of a notion

whose content is Twain, and an idea whose content is the property of being a great

author.[46]

Because notions and ideas are particulars, they can be individuated extremely finely. Though Crimmins and Perry do not state specific individuation conditions for notions and ideas, Crimmins suggests that notions are to be individuated finely enough to account for the phenomenon of recognition failure:

> When an agent perceives an individual and forms beliefs about the object of perception, she may or may not recognize the individual. If she recognizes it, she will connect the perception with a pre-existing notion. . . Failure of recognition involves having multiple notions of a single individual. When I do not recognize you at a distance, I do not connect the notion I form in perception with my stable notion of you. Many of the belief puzzle cases turn on just this phenomenon: an agent has two unconnected notions (often, both stable notions) that happen to be of the same individual.[47]

Thus if it is in some sense *possible* for Odile to fail to recognize Twain on a given occasion, then Odile must at that time instantiate at least two distinct notions of Twain. In order to individuate notions (and presumably ideas) this finely, there must be many different conditions which are *sufficient* for the individuation of cognitive particulars. For instance, Odile may have one notion of Twain which she associates with the name 'Twain', and a distinct notion she associates with the name 'Clemens'. She also may have a notion of Twain which is associated with the sound of his voice, and perhaps a notion which is involved in a particular sighting of Twain. Notions can also be individuated by appeal to the beliefs of which they are constituents: Odile may have one notion which is involved in her belief that Twain wrote *Tom Sawyer*, and another notion which is associated with her belief that Twain wrote *The Adventures of Huckleberry Finn*. And finally, Crimmins and Perry suggest that normal adults typically possess "normal notions" of famous individuals. For example, Odile may have a "normal notion" of Twain, a notion which is associated with a set of stereotypical beliefs about Twain; a normal notion of Twain may be a notion which is

associated with ideas of properties such as being a great author, being the author of *The Adventure of Huckleberry Finn*, and being witty. (In order to have a normal notion, one must have all of the requisite kinds of beliefs; e.g. assuming that the above is what it is to have a normal notion of Twain, if Odile did not have a belief with the content that Twain is a great author, then she would not posses a normal notion of Twain.)

The advantage of identifying the third elements of instances of propositional attitudes with particulars instead of multiply realizable universals is that in doing so Crimmins and Perry almost guarantee that their theory succeed in precluding the epistemological arguments from opacity.[48] Odile assents to (O1), yet dissents from (O2). According to Crimmins and Perry's theory, it follows from this that

$$(K^{**}) \quad BEL^{**}<Odile, P, b>$$

and that

$$(L^{**}) \quad \neg(BEL^{**}<Odile, P, b'>)$$

(where 'P' designates the Russellian proposition that Twain is a great author, and 'b' and 'b'' designate *beliefs*). Since b and b' are mental *tokens* of some kind, there is (almost) no question of their being identical: b and b' are parts, or features of, the particular brain-events which are responsible for Odile's assent to (O1) and dissent from (O2), respectively, and since b and b' occur at different times, there is (almost) no question of their being the same belief.[49] Hence, because Crimmins and Perry identify the mediators of propositional attitudes with particulars, their theory can successfully preclude the epistemological arguments from opacity.

How do Crimmins and Perry propose to account for the cognitive significance of occurrences? The advantage of identifying the mediators of propositional attitudes

with particulars, rather than multiply realizable universals, is that in doing so Crimmins and Perry insulate their theory from refutation via the epistemological arguments from opacity. The disadvantage of identifying the mediators of propositional attitudes with particulars is that, because they are particulars, the mediators cannot account for the cognitive significance of occurrences. It may be instructive here to contrast Crimmins and Perry's theory with Frege's theory. Frege accounts for the cognitive significance of occurrences by appeal to senses, or thoughts, which are multiply realizable. For example, according to Frege's theory, Odile assents to (O1) because (O1) presents a certain thought for Odile, and she dissents from (O2) because this occurrence presents a distinct thought for her. In stating such an explanation of Odile's behavior, the Fregean theorist is implicitly appealing to a general psychological law of the form

> For all subjects X, if conditions C obtain, and X
> grasps a proposition via T_1, then X will assent.

where conditions C probably include facts concerning Odile's "background" beliefs and desires, her abilities, etc., and 'T_1' designates the mode of apprehension (O1) displays for Odile. Note that in order for this law to make sense, T_1 must be a multiply realizable--it must be possible for there to be a number of instances of subjects grasping propositions via T_1. Crimmins and Perry, however, have no recourse to such general laws; i.e. the following cannot serve as a general law of psychology:

> For all subjects X, if conditions C obtain, and X grasps
> a proposition via cognitive particular CP, then X will assent.

The problem is that there cannot be *instances* of subjects grasping propositions via a cognitive particular: if CP is a cognitive particular instantiated by Odile, and CP' is a cognitive particular instantiated by Oscar, then CP is not identical to CP'.

Of course it does not follow from the fact that Crimmins and Perry's theory does nothing in the way of providing an account of the cognitive significance of occurrences, that their theory is incompatible with there being such an account. After arguing that the epistemological arguments from opacity force one to identify the mediators of propositional attitudes with cognitive particulars (though Crimmins and Perry would not put it that way), Crimmins points out that

> It does not follow--and it is not true--that there are no instances of believing that can be usefully thought of as involving useful states. It is no doubt conceivable that there is, or might have been, a certain neural state such that, when any person is in that state, the person believes that it is raining then and there. . . . And certainly there are important internal properties that are shared across different instances of believing.[50]

(I assume that one of the ways in which a *state*--a multiply realizable universal--can be "useful" or "interesting" would involve its playing a role in psychological laws, such as the one stated above.) Crimmins and Perry's theory thus diverges from the theories of Salmon and Richard (and even Frege), in that the entities posited in order to preclude the epistemological arguments from opacity do not themselves serve to account for the cognitive significance of occurrences. Rather, and here I am to some extent interpreting Crimmins, it is various properties shared by cognitive particulars which account for the cognitive significance of occurrences.

The second important respect in which Crimmins and Perry's theory differs from the theories of Salmon and Richard is that Crimmins and Perry do not endorse the Principle of Full Articulation:

> The Principle of Full Articulation: The proposition presented by an occurrence of a declarative sentence is wholly determined by (a) the semantic values--the referents and designations--of the phonetically or orthographically realized terms appearing in the sentence, (b) the logical form of the sentence.

More specifically, Crimmins and Perry deny that the propositions presented by occurrences of attitude ascriptions are fully determined by (i) the semantic values of the phonetically and orthographically realized terms appearing in the ascription, and (ii) the logical form of the ascription. Thus Crimmins and Perry a.e able to maintain that occurrences of

(3) Odile believes that Twain is a great author.

and

(4) Odile believes that Clemens is a great author.

present distinct propositions, even though they have the same logical form and are isomorphicly composed of (phonetically or orthographically realized) terms having the same semantic values. According to Crimmins and Perry's analysis of attitude ascriptions, an occurrence of (3) asserts, roughly, that Odile holds the attitude of belief toward the Russellian proposition that Twain is a great author via a certain token sentence of mentalese--a certain cognitive particular--but Crimmins and Perry deny that this cognitive particular is the semantic value of any of the phonetically or orthographically realized terms appearing in the ascription. Rather Crimmins and Perry maintain that this cognitive particular is somehow provided by the context of utterance, or *tacitly referred to* by the utterance as whole. More specifically, Crimmins and Perry analyze an occurrence of (3) as

(3**) $\exists b$ (BEL** <Odile, P, b> & Con(b) = P &
 b is appropriately composed of N_t and I_{ga})

where 'P' designates the Russellian proposition that Twain is a great author, 'Con( )' designates a function from beliefs to their contents, 'N_t' designates a notion of Twain, and 'I_{ga}' designates an idea of being a great author.[51]   And Crimmins and Perry analyze a relevant occurrence of (4) as

$$(4^{**})\ (\exists b)\ (BEL^{**}\ <Odile, P, b>\ \&\ Con(b) = P\ \&$$
$$b \text{ is appropriately composed of } N_c \text{ and } I_{ga})$$

where 'N_c' designates a notion of Twain which is distinct from N_t; N_c is Odile's "*Clemens* way of thinking of Twain". Thus according to Crimmins and Perry, a relevant occurrence of (3) asserts that Odile holds the BEL** relation toward the Russellian proposition that Twain is a great author via a belief which contains N_t as a constituent--N_t is tacitly referred to by the occurrence of (3) as a whole. The occurrence of (4), on the other hand, asserts that Odile holds the BEL.** relation toward the Russellian proposition that Twain is a great author via a belief which contains N_c as a constituent--N_c is tacitly referred to by the occurrence of (4) as a whole. Consequently, since (3**) and (4**) express distinct propositions, Crimmins and Perry can allow that the occurrence of (3) is true, while the occurrence of (4) is false. In this way Crimmins and Perry endeavor to preclude the semantic arguments from opacity.

Crimmins and Perry's strategy for precluding the semantic arguments from opacity assumes that in making attitude ascriptions ordinary speakers *tacitly refer* to cognitive particulars, and thus cognitive particulars are *unarticulated constituents* of propositions. But how can a constituent of a proposition be *specified* by an occurrence if it is not the semantic value of a term appearing explicitly in the occurrence? Or more generally, how can Crimmins and Perry justify their denial of the Principle of Full Articulation? Crimmins claims that "it is very common in natural languages for a statement to exploit unarticulated constituents," and he goes on to point out that

> We report the weather, for example, as if raining and snowing and
> sleeting and dark of night were properties of times, but they are one and
> all relations between times and places. If I say 'it is raining,' you
> understand me as claiming that it rains at that time at some place the
> context supplies.[52]

Similar evidence for the presence of unarticulated constituents would seem to be provided by sentences containing relative predicates. Consider an occurrence of 'Magic Johnson is tall.' The predicate 'tall' is a relative predicate; in order to know whether or not a particular occurrence of 'Magic Johnson is tall' expresses a truth, we have to have some idea as to who or what Magic is alleged to be taller than. But there is no phonetically or orthographically realized expression which has the appropriate relatum, or relata, as semantic value. Hence it seems that, just as with occurrences of 'It's raining', the appropriate relatum is provided by the context of utterance: in some contexts the missing relatum is something like, *the average professional basketball player*, and in these contexts the sentence is false. But in other contexts the missing relatum is something like *the average person*, and in these contexts the sentence is true. Occurrences of sentences containing quantifiers also seem to provide counterexamples to the Principle of Full Articulation: consider an occurrence of 'Everybody was there'. In order to know whether or not an occurrence of this sentence is true, one must know what the domain of the quantifier is. In some contexts the domain might be restricted to all the members of a particular Philosophy department, while in other contexts the domain might include all living humans. Again, however, there is no phonetically or orthographically realized term appearing explicitly in the occurrence which specifies the relevant domain. Crimmins and Perry suggest that phenomena such as these also serve as counterexamples to the Principle of Full Articulation, and that consequently their denial of this principle in order to account for the truth conditions of attitude ascriptions is not an *ad hoc* maneuver.[53]

3.3.2 Criticism of Crimmins and Perry's Russellian Theory.

I will raise three difficulties for Crimmins and Perry's theory. The first concerns the truth conditions of common belief\desire explanations of behavior, while the second and third concern the alleged phenomenon of *tacit reference*, or of a specific unarticulated constituent's being *provided by a context*. Again, the difficulties encountered by Crimmins and Perry's theory are reminiscent of the Scylla and Charybdis of the Fregean strategy.

My first objection concerns the ability of Crimmins and Perry's theory to account for our untutored intuitions concerning the truth conditions of ordinary belief\desire explanations of behavior. Crimmins seems to think that one of the advantages his theory has over a Russellian theory such as Salmon's is that on his theory such "folk" explanations will *present* or *semantically encode*, rather than merely *pragmatically convey*, truths. That is, Crimmins seems to think that his theory preserves the "literal truth" of explanations such as

(10a) Odile desires Twain come to the party

and

(10b) she believes that if wine is served, then Twain will come to the party,

and that's why

(10c) Odile is going to serve wine.

But if the unarticulated constituents of the proposition(s) presented by an occurrence of this explanation are *particulars*, then it is not at all clear that Crimmins and Perry's theory preserves the literal truth of such explanations. The problem is not that

Crimmins and Perry's theory predicts the wrong truth conditions for each of the sub-sentences of an occurrence of (10a-b-c), rather the difficulty is that Crimmins and Perry's theory does not account for our untutored intuitions with regard to (10a-b-c) as a whole. More specifically, Crimmins and Perry's theory cannot account for why the conjunction of the propositions expressed by occurrences of (10a) and (10b) *predicts*, or *explains*, the proposition expressed by an occurrence of (10c).

It will be useful to again contrast Crimmins and Perry's theory with Frege's theory. Recall that according to Frege's theory the above "folk" explanation is of the form

Fregean Explanation Form.

(a) X desires content C
(b) X believes content (CV(E occurs) $\rightarrow$ C is realized)
Therefore,
(c) X acts so as to make E occur.

('X' in the above formula is a schematic letter open for replacement by names for subjects; 'C' is a schematic letter open for replacement by terms referring to Fregean thoughts; 'E' is a schematic letter open for replacement by expressions designating potential events. 'CV' denotes a function from semantic values to cognitive values; i.e. from propositions to the thoughts which determine them. I here overlook the fact that there could be no such function--for every proposition there are many thoughts which present it.) Consequently, the Fregean theorist can account for the truth of (10a-b-c) by maintaining that an occurrence of (10a-b-c) implicitly appeals to a general psychological law of the form

Fregean Law Form

For all subjects x, if
(a) x desires content C
(b) x believes content (CV(E occurs) $\rightarrow$ C is realized)
then (*ceteris paribus*)
(c) x acts so as to make E occur.

In this way the Fregean theorist can account for truth conditions of an occurrence of (10a-b-c) *as a whole*. The conjunction of the propositions presented by occurrences of (10a) and (10b) predicts, or explains, the proposition presented by an occurrence of (10c) because the propositions presented by occurrences of (10a) and (10b) in conjunction with the appropriate psychological law entail the proposition presented by occurrences of (10c). Thus the Fregean theorist can maintain that an occurrence of (10a-b-c) is true for essentially the reason that an occurrence of 'That wire is copper, and that's why it conducts electricity' is true; both claims are true in virtue of a relevant general covering law.

Crimmins and Perry, however, cannot avail themselves to such an account of the truth conditions of (10a-b-c). More specifically, Crimmins and Perry do analyze (10a-b-c) as being of a form analogous to the Fregean Form:

Crimmins and Perry Explanation Form

(a) X has desire D
(b) X has belief (CP(E occurs) $\rightarrow$ D is realized)
Therefore,
(c) X acts so as to make E occur.

('D' is a schematic letter open for replacement by terms designating cognitive particulars which are *desires*. 'CP' designates a function from semantic values to cognitive particulars which present them. Again, I here overlook the fact that it is not even plausible to suppose that there is such a function--for every proposition there are

many cognitive particulars which present it.) However, because they maintain that

desires and beliefs are cognitive particulars, there can be no general psychological laws

of the form

> ### Crimmins and Perry Law Form
>
> For all subjects x, if
> (a) x has desire D
> (b) x has belief (CP(E occurs) $\rightarrow$ D is realized)
> then (*ceteris paribus*)
> (c) x acts so as to make E occur.

According to Crimmins and Perry's analysis, the sub-sentences of an occurrence of

(10a-b-c) do not assert that Odile instantiates some kind of multiply realizable universal

that is explanatorily relevant to Odile's behavior. As a result, their analysis cannot

account for our untutored intuitions concerning the truth conditions of such "folk"

explanations of behavior. Even if occurrences of (10a), (10b) and (10c) are true on

Crimmins and Perry's analysis, their analysis cannot account for why the propositions

presented by occurrences of (10a) and (10b) *explain*, or *predict*, the proposition

presented by an occurrence of (10c). Consequently, contrary to what is suggested by

Crimmins, it is not all clear that Crimmins' theory preserves the "literal truth" of "folk"

explanations of behavior.[54]

My second objection to Crimmins and Perry's theory concerns the alleged

phenomenon of *tacit reference*, or of a propositional constituents *being provided by a*

context. Consider again an occurrence of 'It's raining'. Suppose I look out my window

and utter this sentence--suppose that I am just telling you in an offhand way that it is

raining. What proposition have I presented? According to Crimmins and Perry "you

understand me as claiming that it rains at that time at some place the context supplies."

But precisely which place is tacitly referred to by my utterance? Can any of the

following candidates plausibly be identified as the proposition presented by my utterance as a whole?

(a) Rains <t, directly outside my window>
(b) Rains <t, my block>
(c) Rains <t, Cambridge>
(d) Rains <t, the greater Boston area>
(e) Rains <t, Eastern Massachusetts>

Note that (a)-(e) all have different truth conditions; e.g. it can be raining on my block, even though it is not raining directly outside my window. Therefore, assuming that this list exhausts the plausible candidates (though it clearly does not), only one of (a)-(e) can be the proposition I asserted. But which one? This question clearly has a false presupposition: In uttering 'It's raining', I intended to assert proposition (a) no more, nor no less, than I intended to assert proposition (b), and thus there is no reason to interpret me as having asserted proposition (a) as opposed to proposition (b), or vise versa. An utterance of 'It's raining' in a context c typically has the same truth conditions as an utterance of 'It's raining here and now' in c, and the phonetically or orthographically realized indexical 'here' no doubt has a place as its semantic value. But it is a mistake to attempt to specify exactly which place is specified by the occurrence of 'here'--this occurrence of 'here' is not coreferential with any nonindexical place description, nor with any nonindexical place name. Therefore to interpret me as having expressed one of (a)-(e) is to alter, and embellish, what I asserted.[55]

My objection to Crimmins and Perry's analysis of propositional attitude ascriptions is that, just as it is implausible to suppose that my utterance of 'it raining' tacitly refers to, e.g., the space directly outside my window, so it is implausible to suppose that occurrences of attitude ascriptions tacitly refer to, and thereby specify, particular notions and ideas. Consider an occurrence of the ascription

(35) She believes that he is a great leader.

where 'he' refers to Clinton, and 'she' refers to Odile. (Suppose that we are at a political rally at which Clinton is giving a speech. We are observing Odile, whom we do not know, and Odile is obviously enjoying Clinton's speech.) What proposition do I assert in making this utterance? Again, consider some plausible alternatives:

(a) $\exists b$ (BEL <Odile, Q, b> & Con(b) = Q &
   b is appropriately composed of N_v and I_{gl})

(where Q is the Russellian proposition that Clinton is a great leader and
N_v is the notion linked to Odile's current *visual* perception of Clinton;
this is the perceptual notion Odile would be employing if she were deaf,
or simply were not listening.)

(b) $\exists b$ (BEL <Odile, Q, b> & Con(b) = Q &
   b is appropriately composed of N_a and I_{gl})

(where N_a is the notion linked to Odile's current *audio* perception of
Clinton; this is the perceptual notion Odile would be employing if she
were blind, or simply were not watching.)

(c) $\exists b$ (BEL <Odile, Q, b> & Con(b) = Q &
   b is appropriately composed of N_g and I_{gl})

(where N_g is the notion involved in Odile's belief that Clinton is the
former Governor of Arkansas; this is the notion Odile would have if she,
for some reason, did not think that former *Governor* Clinton was
identical to *President* Clinton).

(d) $\exists b$ (BEL <Odile, Q, b> & Con(b) = Q &
        b is appropriately composed of N_n and I_{gl})

(where N_n is the "normal notion" involved in stereotypical beliefs about Clinton; this is the "normal notion" Odile would have if she knew that former Governor Clinton won the presidential election).

Note that, just as with the above case involving 'It's raining', all of (a)-(d) present distinct propositions, and they are all plausible candidates (though they are certainly not the only plausible candidates). Which one of these propositions have I asserted with my utterance of (35)? Again, this question has a false presupposition; I intended to assert (a) no more, nor no less, than I intended to express (b). Just as in the case involving 'It's raining', it is a mistake to interpret my utterance as asserting anything as specific as any of (a)-(d). To interpret me as making a claim about some specific notion of Odile's is to embellish, and alter, what I have expressed.[56]

Perry has a potential response to this objection. Perry is willing to concede that *some* occurrences of ascriptions do not specify the notions and ideas employed by the subject of the ascription, but rather merely express constraints upon the notions and ideas employed by the subject.[57] More specifically, Perry allows that my utterance of

(35) She believes that he is a great leader.

may be correctly analyzed as

(35*) $\exists b$ [(BEL <Odile, Q, b> & Con(b) = Q &
        $\exists n$ (C(n) & b is appropriately composed of n and I_{gl})]

'C( )' in the above formula designates some condition, or *constraint*, on Odile's notion of Clinton. For example, C( ) in the above might be one of the following:

C'( ) = is a *normal notion* of Clinton.

C"( ) = is a *perceptual notion* of Clinton.

C'''( ) = is the notion of Clinton Odile is currently employing.

By analyzing my utterance of (35) in this way, Perry avoids the difficulty presented above; (35*) does not presuppose that I *tacitly refer to* a specific notion Odile has of Clinton. Rather (35*) merely requires that I tacitly specify a certain *constraint* on Odile's notion of Clinton.

Reverting to constraints upon notions, however, merely relocates the problem: Just as I do not intend to refer to Odile's visual notion of Clinton any more, or any less, than I intend to refer to her audio notion of Clinton, so I do not intend to "constrain" Odile's notion of Clinton by way of C'( ), any more or any less than I intend to constrain Odile's notion of Clinton by way of C"( ). I simply have no intention to assert a proposition concerning such a specific constraint, and to interpret me as having asserted a proposition concerning such a constraint is again to embellish, and therefore alter, what I have asserted.[58]

My third objection to Crimmins and Perry's theory concerns *tacit reference failure.* If, in making ordinary attitude ascriptions, speakers *tacitly refer to* notions and ideas, then there ought to be cases of tacit reference *failure.* If Crimmins and Perry's analysis is correct, then there ought to be occurrences of attitude ascriptions such that the speaker attempts to specify notions which simply do not exist; Crimmins concedes that such an occurrence should lack a truth value, as the referential presupposition of the occurrence has not been fulfilled. These predictions, however, do not accord with our untutored intuitions. To begin with, Crimmins and Perry's analysis cannot account for our intuitions with regard to ascriptions, and negated ascriptions, which concern

propositional contents that are wholly foreign to the subject. For example, again consider occurrences of

(8) Aristotle believed that Twain was a great author.

and

(8n) Aristotle did not believe that Twain was a great author.

Our untutored intuitions dictate that occurrences of (8) are false, while occurrences of (8n) are true. Aristotle could not possibly have believed that Twain was a great author, as the proposition presented by the that-clause was completely unknown to Aristotle. Thus our untutored intuitions dictate that occurrences of (8) are *false*, and occurrence of (8n) are *true*. Crimmins and Perry's analysis, however, seems to predict that occurrences of both (8) and (8n) lack truth values. If it assumed that an occurrence of (8) attempts to *specify* Aristotle's notion of Twain, then Crimmins and Perry would analyze the occurrence as follows:

(8*) $\exists b$ (BEL <Aristotle, P, b> & Con(b) = P &
         b is appropriately composed of N_t and I_{ga})

(where 'P' designates the Russellian proposition that Twain is a great author, and 'Nt' allegedly designates Aristotle's notion of Twain). The problem is that 'N_t' lacks a referent--Aristotle did not have a notion whose content was Twain--and thus Crimmins and Perry's theory incorrectly predicts that an occurrence of (8) would be neither true nor false. (Crimmins and Perry's analysis of occurrences of (8n) of course faces a similar difficulty.)

Crimmins and Perry's analysis also has difficulty with more familiar sounding ascriptions and negated ascriptions. To borrow an example from Crimmins, suppose we come upon a man who seems to be gazing up at the Washington Monument.[59] Unbeknownst to me, however, the man is blind. Upon seeing the man, I utter,

(36) That yokel believes it's the tallest thing in the world.

Crimmins claims that my utterance attempts to specify the yokel's current visual notion of the Washington Monument. Consequently, since the yokel has no such notion, Crimmins and Perry's theory predicts that--regardless of the yokel's opinions concerning the height of the Washington Monument--my utterance lacks a truth value. But, again, this prediction does not accord with our intuitions. The fact that the man has no current visual notion of the Washington Monument seems to be irrelevant to the truth conditions of my utterance. Consider what the truth conditions of my utterance would be for each of the following situations:

> Situation 1. Suppose that, also unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "It's so good to be here, under the shadow of the tallest thing in the world." (Suppose that he believes this because his tour guide lied to him.)

> Situation 2. Suppose that, again unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "It's so good to be here, under the shadow of the second tallest thing in the world." (Suppose that his tour guide told him a different lie.)

Our intuitions dictate that if situation 1 were the actual state of the world, my utterance of (36) would be *true*, not meaningless. Granted, I am confused. I am wrong in thinking that the exhibited gazing behavior has something to do with the yokel's belief, and thus my *justification* for thinking that the yokel believes it's the tallest thing

in the world is undermined, but these epistemological issues should not be confused with the *truth conditions* of my utterance. What one says can be true even if one thinks that it is true for the wrong reasons. Similarly, our intuitions dictate that if situation 2 were the actual state of the world, my utterance would be *false*. In this case I would be under the same confusion; I misinterpret the exhibited gazing behavior, and thus my justification for thinking that the yokel believes it's the tallest thing in the world is undermined. But, again, this confusion concerning the justification of my claim does not affect the *truth conditions* of my claim. Thus Crimmins and Perry's analysis seems to make the wrong predictions for even quite ordinary claims.[60]

Crimmins and Perry can again invoke the *constraint* analysis of ascriptions in order to preserve our untutored intuitions. They can claim, for example, that an occurrence of (8)--which we would intuitively judge to be false--does not attempt to *specify* Aristotle's notion of Twain, but rather presents a proposition containing a constraint upon Aristotle's alleged notion of Twain. For instance, an intuitively false occurrence of (8) might present that

$$(8^{**}) \quad \exists b \, [(BEL <Aristotle, r, b> \& \, Con(b) = r \, \&$$
$$\exists n \, (C(n) \, \& \, b \text{ is appropriately composed of n and } I_{ga})]$$

(where C( ) is some constraint on Aristotle's alleged notion of Twain. For example, C( ) might be the constraint, *is a notion of Twain*). There is no non referring term in (8**), and thus (8**) accords with our intuitions that an occurrence of (8) would be false. This strategy could also be employed to preserve our intuitions concerning my utterance of (36), and all other cases of alleged tacit reference failure.

There are, however, several serious difficulties with invoking the constraint analysis of ascriptions in order to avoid the difficulties posed by occurrences which allegedly suffer from tacit reference failure. The first difficulty is that we have no way

of determining whether a given ascription merely presents a *constraint on notions*, or if it attempts to *specify the notions* employed by the subject. It is clear from the above that, if it is to preserve our untutored intuitions concerning attitude ascriptions, Crimmins and Perry's theory must analyze at least some occurrences of ascriptions as presenting constraints on notions (and ideas), as opposed to tacitly referring to specific notions. But how is it determined whether a given occurrence of an ascription specifies a notion, or merely specifies a constraint on notions?

Crimmins and Perry recommend the following procedure:

> look at what would happen if the appropriate notions were to fail to
> exist. If the report would then be false, then it is a case of notion
> constraint rather than [specification]; if the report would fail to make a
> claim, then it is a case of (attempted) notion [specification].[61]

This statement of the procedure, however, will not do. The basic idea behind the recommended determination procedure is that in cases in which specification of notions does not seem relevant to the truth conditions of the ascription, then the ascription merely presents a constraint on notions; it does not matter whether we judge the given ascription to be true, or to be false. So the recommended determination procedure is better stated as follows:

> look at what would happen if the appropriate notions were to fail to
> exist. If the report would then be [either true or] false, then it is a case
> of notion constraint rather than [specification]; if the report would fail to
> make a claim, then it is a case of (attempted) notion [specification].

The problem with this procedure for determining whether a given ascription attempts to specify notions, or merely present a constraint on notions, is that for most occurrences of ascriptions, it is not all clear what "the appropriate notions" are. Consider again the case in which we are observing Odile at the Clinton rally. When I utter 'She believes

that he is a great leader' what are the *appropriate* notions? Does my claim concern Odile's perceptual notions, her normal notions, or what? As was pointed out above, there seems to be no answer to this question. But without an answer to this question, the proposition presented by my utterance cannot be determined by Crimmins and Perry's theory.

The second difficulty with invoking the constraint analysis of ascription is that, in reverting to the constraint analysis, Crimmins and Perry's analysis encounters difficulties similar to the semantic arguments from opacity; i.e. in invoking the constraint analysis, Crimmins and Perry's analysis becomes susceptible to the very type of difficulty their theory was designed to avoid. Consider again the case involving the Washington Monument and the yokel. Our untutored intuitions dictate that, if situation 1 is the actual state of the world, my utterance of

(36) That yokel believes it's the tallest thing in the world.

is *true*. Recall situation 1:

> Situation 1. Suppose that, also unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "It's so good to be here, under the shadow of the tallest thing in the world." (Suppose that he believes this because his tour guide lied to him.)

In order to preserve this intuition, my utterance of (36) must present a constraint which is satisfied by one of the yokel's notions of the Washington Monument. What constraint is presented by my utterance? What is C( ) in the following analysis of my utterance?

(36*) $\exists b$ [(BEL \<the yokel, R, b\> & Con(b) = R &

$\exists n$ (C(n) & b is appropriately composed of n and I_{tt})]

(where 'R' designates the Russellian proposition that the Washington Monument is the tallest thing in the world, and 'I_{tt}' designates the yokel's idea of being the tallest thing in the world).

Crimmins informs us that my utterance attempts to express something about a *visual* notion the yokel has of the Washington Monument. But if C( ) is identified with the constraint, *is a visual notion of the Washington Monument*, then analysis (36*) incorrectly predicts that my utterance of (36), relative to situation 1, is *false*. It was stipulated that the yokel is blind, and thus he has no visual notions whatsoever. Therefore if we are to preserve our intuition that relative to situation 1 my utterance of (36) is true, C( ) must be more inclusive than *is a visual notion of the Washington Monument*.

We might try interpreting C( ) in (36*) as the more inclusive constraint, *is a perceptual notion of the Washington Monument*, but this will not do either. Suppose that in situation 1 the yokel has no direct perceptual contact with the Monument whatsoever. That is, suppose that the actual state of the world relative to which I utter (36) is as follows:

> Situation 1'. Suppose that, also unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "It's so good to be here, under the shadow of the tallest thing in the world." Further suppose that he believes this because his tour guide lied to him, and that he has had no direct perceptual contact with the monument whatsoever.

Our intuitions dictate that relative to situation 1', my utterance of (36) is *true*. But if C( ) in (36*) is interpreted as designating the constraint *is a perceptual notion of the*

Washington Monument, then the proposition presented by (36*) is again false, and thus (36*) cannot serve as an analysis of my utterance of (36).

Similarly, interpreting C( ) in (36*) as *is a normal notion of the Washington Monument* will not suffice; again, we can just stipulate into situation 1 that the yokel lacks the requisite stereotypical beliefs about the Washington Monument:

> Situation 1". Suppose that, also unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "It's so good to be here, under the shadow of the tallest thing in the world." Further suppose that he believes this because his tour guide lied to him, and that he has had no direct perceptual contact with the monument whatsoever, and he does not possess the beliefs required for having a *normal notion* of the Washington monument.

If C( ) in (36*) is interpreted as is a normal notion of the Washington Monument, then the proposition presented by (36*) is again *false*, and thus (36*) still cannot serve as an analysis of my utterance of (36).

Even interpreting C( ) as something as inclusive as the constraint, *is a notion of the Washington Monument which is a constituent in the yokel's occurrent belief that he is standing under that thing, (where 'that thing' refers to the Washington Monument)* will not do. Again, situation 1 can merely be stipulated to be such that we intuit that my utterance of (36) would be *true* relative to situation 1, though the analysis presented by (36*) would be false:

> Situation 1'". Suppose that, also unbeknownst to me, moments before I uttered (36) the yokel had sincerely said to you, "The Washington monument is the tallest thing in the world." Further suppose that he believes this because his friend guide lied to him, and that he has had no direct perceptual contact with the monument whatsoever, and he does not possess the beliefs required for having a normal notion of the Washington monument, and he does not know where he is, though his mental disposition is such that if his surroundings where described to him, he would declare, "I am beside the Washington Monument, which is the tallest thing in the world".

Our untutored intuitions dictate that if I were to utter (36) relative to situation I''', then my utterance would be *true*. The analysis presented by (36*), where C( ) is interpreted as designating the constraint, *is a notion of the Washington Monument which is a constituent in the yokel's occurrent belief that he is standing under that thing, (where 'that thing' refers to the Washington Monument)*, would be false (or meaningless) relative to situation I'', as in this situation the yokel has no occurrent belief that he is standing under *that thing*.

The difficulty manifesting itself here arises because our intuitions dictate that as long as the yokel is disposed to sincerely utter things like, "The Washington Monument is the tallest thing in the world," my utterance is true; the facts concerning how the yokel is epistemologically linked to the Washington Monument are irrelevant to the truth conditions of my utterance. Consequently, if the analysis presented by (36*) is to preserve these intuitions, C( ) in (36*) must be interpreted so that no matter what notion the yokel employs in his belief that the Washington Monument is the tallest thing in the world--no matter how the yokel is epistemologically linked to the Washington Monument--this notion satisfies C( ). As a result, C( ) must be interpreted as an extremely inclusive constraint.

The problem is that if C( ) is allowed to be extremely inclusive, then Crimmins and Perry's theory becomes susceptible to difficulties similar to the semantic arguments from opacity, and thus fails to preserve the legitimacy of our ordinary attitude ascribing practices. For example, suppose that C( ) is the inclusive constraint, *is a notion of the Washington Monument*. Further suppose that the actual state of the world is as follows:

Situation 3. The yokel, for whatever reason, thinks that there are *two* Washington Monuments, one of which is the tallest thing in the world. (Thus there is some notion of the Washington Monument which is a constituent of a belief that has Russellian proposition R as its content. Let us call this notion, 'N_{tm}'.) Furthermore, let us suppose that the yokel thinks the Monument he is standing next to when I utter (36) is not the tall monument, but rather is the shorter monument. (Thus there is another notion of the Washington monument which is not a constituent of a belief that has Russellian proposition R as its content. Let us call this notion, 'N_{sm}'. Lastly, if the yokel were presented with an occurrence of 'It's the tallest thing in the world," he would dissent.

Our naive intuitions dictate that, if situation 3 were to obtain, my utterance of

(36) That yokel believes that it's the tallest thing in the world.

relative to situation 3 would be *false*, while my utterance of

(36n) That yokel does not believe that it's the tallest thing in the world.

relative to Situation 3 would be *true*. (Note that, if necessary, further stipulations can be added to Situation 3 to solidify these intuitions.) Crimmins and Perry, however, would analyze my utterances of (36) and (36n) as follows:

(36*) $\exists b$ [(BEL <the yokel, R, b> & Con(b) = R &
           $\exists n$ (C(n) & b is appropriately composed of n and I_{tt})]

(36n*) $\neg(\exists b$ [(BEL <the yokel, R, b> & Con(b) = R &
           $\exists n$ (C(n) & b is appropriately composed of n and I_{tt})])

(where C( ) in (36*) and (36n*) is the inclusive constraint is a notion of the Washington Monument).

The problem of course is that these analyses incorrectly predict that my utterance of (36) relative to Situation 3 is *true*, while my utterance of (36n) relative to Situation 3 is *false*. The constraint *is a notion of the Washington Monument* is so inclusive that it is satisfied by both N_{tm} and N_{ts}. Consequently, relative to Situation 3 (36*) presents a truth--there is some notion of the Washington Monument, viz. N_{tm}, which is a constituent of a belief of the yokel's which has R as its content. And for similar reasons relative to Situation 3 (36n*) presents a falsehood. Thus (36*) and (36n*) do not accord with our intuitions concerning the occurrences of (36) and (36n), respectively, and consequently Crimmins and Perry's theory fails to preserve the legitimacy our ordinary attitude ascribing practices.

This difficulty for Crimmins and Perry's theory is reminiscent of the fundamental difficulty facing the Fregean strategy. In order to preclude the arguments from opacity, the Fregean theorist is compelled to individuate cognitive values extremely finely, but in doing so he undermines his account of attitude ascription; if cognitive values are individuated finely enough to steer clear of Scylla, the Fregean theorist cannot preserve the legitimacy of our ordinary attitude ascribing practices. Analogously, Crimmins argues that the cognitive particulars--the mediators of propositional attitudes--must be identified with particulars, as opposed to multiply realizable universals, because "we must rule out the possibility of two instances of believing, or one of believing and one of not believing, involving the same agent, time, and [mediator]."[62] In other words, Crimmins thinks that the arguments from opacity cannot be successfully precluded if the *mediators* of propositional attitudes are any kind of multiply realizable entity; i.e. Crimmins maintains that the only way of ensuring that the mediators are individuated finely enough to preclude the arguments from opacity is to identify them with neurological particulars, where is impossible for different subjects to instantiate, or "grasp," the same cognitive particular. However, in order to avoid the problems associated with (tacit) reference failure and thus preserve

the legitimacy of our ordinary attitude ascribing practices, Crimmins and Perry must invoke the constraint analysis of attitude ascriptions. But, if the constraint analysis is invoked, then Crimmins and Perry's theory of ascriptions is roughly equivalent to a theory which identifies the mediators of propositional attitudes with very "course grained" multiply realizable universals, and thus the theory encounters difficulties similar to the arguments from opacity.

3.4 Conclusion: The Lessons for Russellian Theorists.

Let us review. Salmon's Russellian Theory differs from The Naive Russellian Theory in that Salmon abandons the binary analysis of propositional attitudes. Salmon's theory attempts to preclude the epistemological arguments from opacity and account for the cognitive significance of occurrences by positing "modes of apprehension" and analyzing propositional attitudes as ternary relations between agents, Russellian propositions, and appropriately individuated "modes of apprehension." It was argued, however, that Salmon does not individuate "modes of apprehension" finely enough--his theory does not satisfy the Individuation Constraint. As a result Salmon's theory can be refuted by epistemological arguments from opacity, and "modes of apprehension" cannot account for the cognitive significance of occurrences. Furthermore, Salmon's theory cannot avoid the semantic arguments from opacity because it endorses, in addition to the principles of Semantic Innocence and Direct Reference, (i) the Principle of Full Articulation, (ii) the Principle of Propositional Truth, and (iii) a straightforward binary analysis of attitude verbs. As a result, Salmon's Theory cannot distinguish the truth conditions of occurrences of

(3) Odile believes that Twain is a great author.

from the truth conditions of occurrences of

(4) Odile believes that Clemens is a great author.

These sentences are isomorphicly constructed out terms which have identical semantic values, and thus occurrences of these sentences must present the same proposition. Consequently, by the Principle of Propositional Truth, occurrences of these sentences must have the same truth conditions.

Richard's theory can be understood as an attempt to improve upon Salmon's theory. With regard to the epistemological arguments from opacity and the problem of the cognitive significance of occurrences, Richard's theory basically follows Salmon's theory except that where Salmon posits "modes of apprehension," Richard posits "representational types." It was argued, however, that Richard's means of individuating representational types violates the Independence Constraint; in the end representational types are individuated by appeal to instances of the phenomenon of opacity itself. As a result representational types cannot account for the cognitive significance of occurrences and Richard's method for precluding the epistemological arguments from opacity is *ad hoc*. With regard to the semantic arguments from opacity, Richard's theory differs from Salmon's theory in that Richard abandons a straightforward binary analysis of propositional attitude verbs in favor of an analysis according to which attitude verbs are complicated indexicals. If Richard's analysis of attitude verbs were correct, then the truth conditions of (many) occurrences of (3) and (4) could be distinguished without abandoning the principles of Propositional Truth and Full Articulation. Against Richard's theory, however, it was argued that Richard's analysis of attitude verbs is incorrect; propositional attitude verbs are not indexical in the way required by Richard's theory. Moreover, Richard's theory of attitude ascriptions does not preserve the legitimacy of our ordinary attitude ascribing practices: Richard's

theory requires ordinary speakers to possess knowledge and abilities which they typically do not possess.

Finally, Crimmins and Perry's theory differs from the theories of Salmon and Richard in two important ways: First, Crimmins and Perry are seemingly able to preclude the epistemological arguments from opacity once and for all because Crimmins and Perry maintain that the mediators of propositional attitudes are *cognitive particulars* rather than multiply realizable universals. The drawback of identifying the third elements of instances of propositional attitudes with neurological tokens rather than multiply realizable entities is that the mediators themselves cannot account for the cognitive significance of occurrences. Thus Crimmins and Perry's theory does not in and of itself offer any kind of solution to the problem of the cognitive significance of occurrences. With regard to the semantic arguments from opacity, Crimmins and Perry's theory differs from the theories of Salmon and Richard (and also The Naive Russellian Theory) in that Crimmins and Perry deny the Principle of Full Articulation: Crimmins and Perry allow that the proposition expressed by an occurrence may contain *unarticulated constituents* which are *tacitly referred to* by the occurrence as a whole. Thus, by appeal to unarticulated constituents, the propositions presented by occurrences of (3) and (4) are distinguished, and therefore, by the Principle of Propositional Truth, the truth conditions of these occurrences are distinguished.

As was shown above, however, Crimmins and Perry's theory of attitude ascription faces several difficulties: First, analyzing occurrences of attitude ascriptions as "tacitly referring to" unarticulated constituents, or even as presenting constraints on such constituents, seems to embellish, and therefore alter, what is actually asserted by occurrences of ascriptions; in most cases there is just no saying what the alleged unarticulated constituents, or constraints, are. Second, Crimmins and Perry's theory encounters difficulties associated with the alleged phenomenon of tacit reference

failure, and consequently Crimmins and Perry's theory also fails to preserve the legitimacy of our ordinary attitude ascribing practices.

What are the lessons for Russellian theorists? How can the Russellian theories examined in this chapter be improved upon? It is, I hope, relatively clear that all of the Russellian theories examined here run afoul of the fundamental difficulty with the Fregean strategy: All of the Russellian theories considered in this chapter attempt to steer clear of Scylla--the problem of precluding the epistemological arguments from opacity and accounting for the cognitive significance of occurrences--by positing finely individuated entities to serve as the *mediators* of propositional attitudes. But in steering away from Scylla in this way, they steer directly for Charybdis, the problem of preserving the legitimacy of our ordinary attitude ascribing practices. It should be noticed that all three of the Russellian theories at least implicitly endorse two crucial assumptions: First, all of the theorists considered in this chapter at least implicitly endorse the Principle of Propositional Truth; they all assume, for example, that the truth conditions of *occurrences* of (3) and (4) differ only if the *propositions presented by these occurrences* are distinct. Second, they assume that the propositions presented by these occurrences can be distinct only through the presence of the finely individuated *mediators* of belief; they all assume, for example, that if the propositions presented by occurrences of (3) and (4) are to be distinguished, it must be by appeal to the distinct *mediators* utilized by Odile. (Salmon simply denies that occurrences of (3) and (4) have distinct truth conditions, because his theory has no way of incorporating "modes of apprehension" into the propositions presented by occurrences of (3) and (4). Richard does not incorporate the mediators of belief directly into the propositions presented by occurrences of (3) and (4), but rather "smuggles in" reference to the mediators of belief by way of his complicated analysis of attitude verbs; what distinguishes the semantic values of 'believes' are restrictions, and restrictions contain sets of mediators. And finally Crimmins and Perry simply claim that the mediators of

belief are *unarticulated constituents* of the propositions presented by occurrences of (3) and (4).) It is because of these two crucial assumptions that the Russellian theorists considered in this chapter get caught between Scylla and Charybdis: the three theories considered here fail to preserve the legitimacy of our ordinary attitude ascribing practices because they all assume that, e.g., the truth conditions of occurrences of (3) and (4) can be distinguished only if these occurrences somehow represent the distinct mediators utilized by Odile. This, however, runs counter to the fact that our ordinary attitude ascribing practices do not require any knowledge of, or abilities to refer to, anything like the posited mediators of propositional attitudes.

I suggest then that the lesson for the Russellian theorist is this: both assumptions in the above paragraph must be rejected. First, the Principle of Propositional Truth should be rejected. Note that, though the Principle of Propositional Truth is at least implicitly endorsed by all three of the Russellian theorists, no argument is offered in support of the principle. Moreover, I suggest that there is good to reject the principle. Given the arguments in support of the principles of Direct Reference and Semantic Innocence, and the intuitive plausibility of the binary analysis of attitude verbs and the Principle of Full Articulation, it seems that the semantic arguments from opacity themselves count as arguments against the Principle of Propositional Truth: Our untutored intuitions to the effect that an occurrence of

(3) Odile believes that Twain is a great author.

may be true in c, while an occurrence of

(4) Odile believes that Clemens is a great author.

may, at the same time, be false in c, should be interpreted *not* as evidence that such occurrences do not present the Russellian propositions they seem to present, but rather as evidence against the Principle of Propositional Truth. That is, instead of preserving the Principle of Propositional Truth at the cost of either our untutored intuitions concerning truth conditions or the straightforward Russellian analysis of attitude ascriptions, I propose that the Principle of Propositional Truth be rejected in order to preserve our intuitions concerning truth conditions and the straightforward Russellian analysis of attitude ascriptions.

Second, the assumption that the mediators of belief, whatever they may be, must be somehow represented by a true attitude ascription should be abandoned. Note that if this second assumption is denied, then the Russellian theorist will no longer be forced to steer between Scylla and Charybdis (or at least the channel between them will be significantly broadened). If occurrences of (3) and (4) are not interpreted as saying something about the complex epistemological and psychological facts which account for Odile's behavior, then whatever epistemological theory we employ in order to preclude the epistemological arguments from opacity and/or account for the cognitive significance of occurrences will have no impact, or at least very little impact, on our theory of propositional attitude ascriptions. I.e. if it is denied that ordinary attitude ascriptions are in any sense about the complex psychological facts and conditions which underlie, and are required for our having beliefs (desires, etc.), then these complicated conditions can be individuated as finely as one likes without thereby threatening the legitimacy of our ordinary attitude ascribing practices. This is not to say, however, that these complex psychological facts are not relevant to, or even necessary for, the truth of ordinary attitude ascriptions. For example, an occurrence of

(3) Odile believes that Twain is a great author.

is no doubt true only if certain complicated neurological and psychological facts obtain. But it does not follow that this occurrence presents a proposition concerning these complicated facts.

Perhaps an analogy can clarify this point. Consider an occurrence of

(37) Twain lived.

The truth of this occurrence is of course dependent upon all kinds of complicated psychological facts concerning what it is to be alive. Suppose that humans are alive if and only if they instantiate some complicated bio-chemical property involving DNA and beta-waves. Under this supposition, an occurrence of (37) is true only if Twain instantiated this complex property. It would be a mistake, however, to suppose that an occurrence of (37) *says something about* DNA and beta-waves; it would be a mistake to suppose that understanding an occurrence of (37), knowing the truth conditions of an occurrence of (37), requires one to have some knowledge of DNA and beta-waves. An occurrence of (37) is not about DNA and beta-waves and thus understanding such an occurrence requires no knowledge of these specialized facts. Similarly there can be little doubt that an occurrence of

(3) Odile believes that Twain is a great author.

is true only if Odile instantiates some complicated psychological, or neuro-chemical, property. But it would be a mistake to conclude from this that knowing the truth conditions of an occurrence of (3) requires one to have detailed knowledge of this obscure psychological property.

Chapter 3 Notes

[1] Salmon presents his theory in Salmon 1986, and Salmon 1989. Richard presents his theory in Richard 1990, and Richard 1992. Crimmins and Perry present their theory in Crimmins and Perry 1989. The theory is presented in more detail in Crimmins 1992a, and is also sketched in Crimmins 1992b.

[2] It is somewhat tendentious of me to interpret Richard's theory as respecting the Principle of Semantic Innocence. Though Richard's theory can be interpreted so that it respects this principle, Richard does not claim to endorse the principle. In the conclusion of his book, Richard states, "With Frege and against Russell, I do not think an attitude is characterized simply in terms of objects and properties. And like Frege, I see this as requiring a sort of reference shift when an expression is embedded," (Richard 1990, p. 265).

[3] Salmon 1986, p. 111. (The underlining is mine.)

[4] Salmon's analysis of attitude verbs can perhaps be more perspicuously represented in terms of "lambda notation." Salmon can be understood as maintaining that 'believes' ('desires' etc.) designates

$$\lambda x \lambda p \ [\exists m \ (BEL <x, p, m>)].$$

[5] That "modes of apprehension" are to account for the cognitive significance of occurrences is made manifest in the following passage concerning the fabled ancient astronomer who assents to an occurrence of 'Hesperus is Hesperus,' yet dissents from an occurrence of 'Hesperus is Phosphorus':

> When he reads and understands the sentence, 'Hesperus is Phosphorus', he takes the proposition thereby encoded in a way different from the way in which he takes this same proposition when he reads and understands the sentence 'Hesperus is Hesperus'. He grasps the very same proposition in two different ways, by means of two different guises, and he takes this single proposition to be two different propositions. When he takes it as a singular proposition of self-identity between the first heavenly body sometimes visible in such-and-such location at dusk and itself, he unhesitatingly assents inwardly to it. When he takes it as a singular proposition identifying the first heavenly body sometimes visible in such-and-such a location with the last heavenly body visible in so-and-so location at dawn, he has no inclination to assent inwardly to it, and may even dissent from it. His verbal assent and his refraining from verbal assent with respect to the two sentences are merely the outward manifestations of his inward dispositions relative to the ways he takes the proposition encoded by the two sentences, (Salmon 1986, p. 113).

6 See Salmon 1986, p. 119-20, 126.

7 Salmon 1986, p. 107-8.

8 I assume that occurrences of sentences suffering from reference failure are neither true nor false. I think this view of occurrences containing empty terms is correct, though the *correctness* of this view is not presupposed by the arguments I present against the subject oriented proposal; i.e. my arguments do not presuppose that on the final analysis occurrences of sentences containing empty terms will turn out to neither true nor false. Some of my arguments against the subject oriented proposal are of the following form: "The subject oriented proposal predicts an instance of reference failure for a particular occurrence, and therefore, if this proposal is correct, our intuitions should dictate that this occurrence is neither true nor false. But our intuitions do not dictate this, and consequently the subject oriented proposal is false." Thus all that is presupposed by my arguments is that our untutored intuitions, regardless of whether or not these intuitions are correct, dictate that sentences containing empty terms are neither true nor false. This assumption is, I think, indubitable.

9 In Salmon 1989, p. 248-50, Salmon summarizes the three reasons he advances in *Frege's Puzzle*.

10 Salmon 1989, p. 249. (Note the implicit appeal to the Principle of Propositional Truth in this passage.)

11 Salmon 1986, p. 116-117. (Again note the emphasized portion of this passage, in which Salmon commits himself to the Principle of Propositional Truth.)

12 In the following passage Salmon's third reason and his commitment to the Principle of Propositional truth is further elucidated:

> Since the form of words 'Hesperus is Phosphorus' is considerably richer
> in pragmatic impartations than other expressions having the same
> semantic information content (e.g. 'Hesperus is Hesperus'), if one is not
> careful one cannot help but mistake the 'that'-clause as referring to this
> somewhat richer information--information which [the subject] does not
> believe. . . . Utterances of the locution ⌐a believes that S⌐ may even
> typically involve a Gricean implicature to the effect that the person
> referred to believes the information that is typically pragmatically
> imparted by utterances of S. *Even so, that is not part of the literal
> content of the belief attribution* [emphasis added]. The general masses,
> and most philosophers, are not sufficiently aware of the effect that an
> implicature of this kind would have on ordinary usage. It is no
> embarrassment to [Salmon's theory] that ordinary speakers typically deny
> literally true belief attributions . . . when these attributions involve a 'that'-
> clause whose utterance typically pragmatically imparts information which
> the speaker recognizes not to be among the beliefs . . . of the subject of
> the attribution, (Salmon 1986, p. 84-5).

13 Salmon 1986, p. 108-9, and throughout Chapter 8. (Recall that "the mode of acquaintance by which one is familiar with a particular object is part of the mode of apprehension by which one grasps a singular proposition involving that object," Salmon 1986, p. 108.)

14 In a very important footnote, Salmon states,

> one cannot require . . . that . . . ways of taking objects are rich enough by
> themselves to determine the object so taken, without the assistance of
> extra-mental, contextual factors. Presumably, twin agents who are
> molecule-for- molecule duplicates, and whose brains are in exactly the
> same configuration down to the finest detail, may encounter different
> (though duplicate) objects, taking them in the very same way. Likewise,
> a single agent might be artificially induced through brain manipulations
> into taking different objects the same way, (Salmon 1989, p. 275).

Salmon *almost* concedes here that his theory is susceptible to refutation via epistemological arguments from opacity: if a subject A can "take two objects in the same way," then surely it is nomologically possible to convince such a subject that he is taking, or has taken, two objects in the same way. But if we could convince this subject of this, then we could also lie to him, and convince him that he has "taken two objects in one way," where he has really "taken one object in one way." If Salmon concedes this much, then he concedes that his theory can be refuted by epistemological arguments

from opacity. (Also note this passage implies that modes of apprehension are universals.)

[15] Ordinary speakers do not intuit that all occurrences of (4) are false--in some "insensitive" contexts they will judge an occurrence of (4) to be true, and in other contexts they will be hesitant to judge it to be either true or false.

[16] This statement is too strong. As will be shown in the next section, Richard attempts to maintain the principles of Direct Reference, Semantic Innocence, Full Articulation and Propositional Truth by construing propositional attitude verbs as complicated *indexicals*, and thereby distinguishing the propositions presented by occurrences of (3) and (4).

[17] I interpret Richard as maintaining that propositional attitudes are ternary relations, though it also plausible to interpret Richard as following Salmon and maintaining that propositional attitudes are binary relations: Recall that Salmon maintains that occurrences of 'believes' designate existential generalizations of instances of the ternary BEL relation; i.e. the semantic value of 'believes' is the following binary relation:

$$\lambda x \, \lambda p \, (\exists m \, (BEL <x, p, m>)) \, (\_\_,\_\_)$$

(where m ranges over modes of apprehension). Similarly, Richard can be understood as claiming that the attitude of belief is the following binary relation:

$$\lambda x \, \lambda p \, (\exists \sigma \, (BEL^* <x, p \, \sigma>))(\_\_,\_\_)$$

If Richard where interpreted in this way, then, just as Salmon must reject the Semantic Principle of Dissent, Richard would be forced to reject the Epistemological Principle of Dissent. Though it is plausible to interpret Richard in this way, I think it is more in the spirit of his theory to interpret him as following Salmon and claiming that propostional attitudes are ternary relations. (Note that in Chapter 4 I will suggest that something like this alternative interpretation of Richard is the correct analysis of propositional attitudes.)

[18] Richard 1990, p. 4.

[19] Richard 1990, pp. 3, 41, 186.

[20] Richard never explicitly endorses these principles: In the examples Richard considers, he simply states what tokens he takes to be written on the subject's belief blackboard, without considering the evidence required to support such statements. Nonetheless, it is relatively clear from what Richard does say that he is committed to these principles. That Richard is committed to this version of the Epistemological Principle of Assent is relatively clear, though a brief argument illustrating why Richard is committed to this version of the Epistemological Principle of Dissent is called for: The reason that Richard is committed to this principle of dissent is that not endorsing

such a principle is tauntamount to abandoning our untutored intuitions concerning the truth conditions of occurrences, the very intuitions Richard's theory is designed to preserve. Suppose that Richard denies this principle of dissent, yet maintains this principle of assent. Further suppose that, as is compatible with this principle of assent, Odile encounters (and understands etc.) an occurrence of Σ, and dissents from it, even though she has a token of Σ written on her belief blackboard. The problem is that, as will shown be when Richard's view of ascriptions is explicated, it is sufficient for the truth of an ascription of the form ⌜Odile believes that Σ⌝ in any context c, that Odile have a token of Σ written on her belief blackboard. Consequently, an ascription of the form ⌜Odile believes that Σ⌝ is true in any context, even though Odile dissents from an occurrence of Σ. This result is incompatible with our untutored intuitions concerning attitude ascriptions.

[21] That sentences of mentalese account for the cognitive significance of occurrences is suggested by Richard's justification of folk explanations such as, "Randi waved because Randi wanted Hesperus to rise, and he thought that if there was waving, Hesperus would rise." In the process of justifying such explanations, Richard states,

> It is plausible to think that, if Randi accepts a sentence of the form
>
> $$W \rightarrow Rh$$
>
> where W is a sentence with the truth condition of 'there is waving', Rh is a sentence in which 'h' names Hesperus, 'R' picks out the property of rising and he "desires-true" a token of Rh, this will make, or tend to make, him wave, (Richard 1990, p. 174).

(For our purposes, accepting a sentence is equivalent to having a token of that sentence written on one's belief blackboard.)

[22] Richard 1990, p. 2.

[23] Richard 1990, p. 60.

[24] Richard 1990, p. 4.

[25] Richard 1990, p. 136.

[26] Richard's theory thus shares with Salmon's theory the rather odd feature that occurrences of ascriptions of the form, ⌜N believes that Σ⌝ do not assert that the referent of N holds *the* epistemological attitude of belief toward the content expressed by Σ. Richard thus seems committed to the view that attitude ascriptions are not strictly speaking about propositional attitudes. (Recall that a subject holds the attitude of belief toward a Russellian proposition p if and only if he has a mental sentence written on his

belief blackboard which presents p.) Given this it is difficult to see how Richard's theory could be a part of "folk psychology" as Richard claims it is.

27 A corresponding semantic argument from opacity can be constructed from this case. If Peter assents to an occurrence of (17), and dissents from another occurrence of (17), it follows from Richard's versions of the epistemological principles of assent and dissent that a token of (17) is both written on, and not written on, Peter's belief blackboard. Consequently, Peter both has and does not have the RAM

RAM (17):

<<IS A TALENTED MUSICIAN, being a musician>, <PADEREWSKI, Paderewski>>

in his RS. Consequently, in many contexts both the ascription

(17') Peter believes that Paderewski is a talented musician.

and the (negative) ascription

(17'n) Peter does not believe that Paderewski is a talented musician.

will be true. For example, if a context c determines no restrictions, then (17') is true in c because RAM (17) is in Peter's RS, but (17'n) is true in c because RAM (17) is not in Peter's RS. (I assume that there is no other RAM in Peter's RS which the that-clause of (17'n) could represent.) In this way Kripke's 'Paderewski' cases can also be used to generate semantic arguments from opacity against Richard's theory.

28 Actually, Richard never considers the epistemological arguments from opacity, and thus I am to some extent constructing what his response to these arguments would be. Because Richard does not distinguish the epistemological arguments from opacity from the semantic arguments from opacity, he is only concerned with precluding the semantic arguments from opacity. Thus he is only concerned to distinguish the RAM associated with Peter's assent to the occurrence of (17) from the RAM associated with Peter's dissent from the other occurrence of (17). This much can be accomplished by distinguishing the linguistic elements of these RAMs and Richard attempts to do this by identifying the linguistic elements with *representational types*. This same general idea can be applied to the epistemological arguments from opacity, and I have thus taken the liberty of augmenting Richard's theory accordingly.

[29] The idea of a causal or historical chain is taken from the causal theory, or "picture," of reference. The idea, first proposed in Kripke 1972, is that there is an initial "dubbing event" in which a token τ is somehow used to refer to an individual for the first time. A second speaker may now use another token τ' with the intention of referring to whatever the first speaker referred to with his use of τ. After hearing the second speaker, a third speaker may use yet another token τ'' with the intention of referring to whatever the second speaker referred to with his use of τ', and so on. (Of course τ, τ' and τ'' will be orthographically and/or phonetically similar.) In this way a sort of historical tree of transmission is constructed; some tokens of type τ will be members, or nodes, of this tree, and others will not be.

[30] Richard 1990, p. 184.

[31] Richard's response to the corresponding semantic argument from opacity (See note 27 above.) is as follows: Because representational types are now the linguistic elements of annotations, it follows from Peter's assent to token t_σ that the following RAM is in Peter's RS:

RAM t_σ:

<<PAD$_1$, Paderewski>, <IS A TALENTED MUSICIAN, is a talented musician>>

where 'PAD$_1$' designates the representational type of the token of 'Paderewski' appearing in the token t_σ, and the mentalese correlate token of t_σ which is written on Peter's belief blackboard. Moreover, it follows from Peter's dissent from token $t_{\sigma'}$ that the following RAM is not in Peter's RS:

RAM $t_{\sigma'}$:

<<PAD$_2$, Paderewski>, <IS A TALENTED MUSICIAN, being a talented musician>>

where 'PAD$_2$' designates the representational type of the token of 'Paderewski' appearing in $t_{\sigma'}$ and the mentalese correlate of $t_{\sigma'}$ which is not written on Peter's belief blackboard.

Once representational types are introduced as the linguistic elements of annotations, finer grained restrictions can be invoked and Richard can avoid the semantic argument from opacity against his theory. Suppose that, as is described above, Peter has RAM t_σ in his RS, but does not have RAM $t_{\sigma'}$ in his RS. It follows that in all contexts the ascriptions

(17') Peter believes that Paderewski is a talented musician.

and

(17'n) Peter does not believe that Paderweski is a talented musician.

cannot both be true. Assume, for *reductio*, that there is some context c in which both (17') and (17'n) are true. In order for (17') to be true in c, there must be a correlation function that maps the annotation

<'Paderewski', Paderewski>

to the annotation

<Pad$_1$, Paderewski>

which obeys all the restrictions determined in c. (I am assuming that Peter refers to Paderewski only via tokens of 'Paderewski'.) But in order for (17'n) to be true in c, there must be no correlation functions which obey the restrictions determined by c and map the annotation

<'Paderewski', Paderewski>

to an annotation occuring in a RAM in Peter's RS. Obviously there can be no such context. Hence by individuating the linguistic elements of the RAMs in a subject's RS as finely as *representational types*, Richard is able to preclude the semantic argument from opacity involving Peter and 'Paderweski'.

[32] In conversation Richard has suggested that the "interior condition" should not be interpreted as a criterion for the identity of representational types, but rather as a heurestic device for determining whether representational type A is identical to representational type B; i.e. satisfying the exterior and interior condtions is not sufficient for the identity of A and B, though failing to satisfy either of these conditions is sufficient for the distinctness of A and B. In the end, representational types will be individuated by some kind of neuro-chemical property. (I should point out that Richard concedes that his explication of the interior condition is not clear.)

[33] I think it is relatively clear how Odile's utterance of, 'I like Twain, but I'm not sure about Clemens' can be construed as an instance of the phenomenon of opacity; the phenomenon of opacity could be redefined so as to include such utterances as instances of opacity. I do not, however, think that this is a serious problem, and thus I will not undertake the task here.

[34] If Richard allows, as seems plausible, that occurrences of infinitive propositional attitude verbals designate the epistemological attitudes, rather than one of the many relations which serve as designatums, then there may be eternal sentences which contain propositional attitude verbs. For example, 'To understand life is a worthy goal' seems to be an eternal sentence which expresses something about the epistemological attitude, or state, of understanding.

35 Another way in which many (perhaps all) belief ascriptions are context sensitive concerns *degrees* of belief; there is no absolute criterion dictating how firm a subject's convictions must be in order for a belief ascription about that subject to be true. Consequently it seems plausible to suppose that the degree of belief required for the truth of an ascription varies from context to context. For example, in a context in which I am conversing with a fellow philosopher, my utterance of 'I believe that God exists' might be true, while in another context in which I am conversing with a Priest, an occurrence of this same sentence might be false, even though my convictions concerning God's existence do not change from context to context. This point also applies to other attitude verbs, and thus there is undoubtedly a sense in which propositional attitudes are context dependent, but this is not the kind of context dependency required by Richard's theory.

36 For example, Richard states, "belief and other propositional attitudes are mediated relations," (Richard 1990, p. 38).

37 Richard briefly considers this problem and he suggests a means of avoiding it. (See Richard 1990, pp. 244-6.) Richard suggests that the Russellian component of annotations containing 'believes', or BELIEVES, ('doubts' or DOUBTS etc.) be identified with the character, or "constant meaning," of 'believes', as opposed to one of the many possible semantic values of 'believes'. It is not clear, to me at any rate, what this proposed amendment amounts to: Is Richard suggesting (a) that the semantic value of an embedded occurrence of 'believes' is *the character* of believes? Do the characters of attitude verbs serve as the *oblique designations* of occurrences of the verbs? Or is the suggestion merely (b) that the RAM determined by a sentence containing a propositional attitude verb does not contain the semantic value of the verb as a Russellian constituent, even though the semantic value of the term is still the relation determined by the character? Crimmins, who criticizes Richard on this point, interprets Richard as maintaining (a), and in responding to this criticism Richard does not correct Crimmins' interpretation. (See note 3 in Crimmins 1992b.) It is not clear, however, that interpretation (a) of Richard's proposal even makes sense: If embedded occurrences of 'believes' designate the character of 'believes', then this character must, in some contexts, determine *itself* as semantic value. But it is not clear that this is coherent: A character is a function which, given a context, determines a semantic value. In set theoretic terms, a character is a set of ordered pairs, the first elements of which are contexts, and the second elements of which are sets. If, given some contexts, the character of 'believes' determines itself as semantic value, then the character of believes contains an ordered pair whose second element is itself. But this of course violates the axioms of standard set theory. Interpretation (b), on the other hand, is also problematic: If the that-clause of an ascription determines a RAM such that the Russellian constituent corresponding to the embedded attitude verb is the character of that verb, and truth is defined in terms of RAMs, then the relation designated by the embedded verb makes no contribution toward the truth conditions of the ascription. Under interpretation (b), the semantic value designated by an embedded occurrence of 'believes' ('desires' etc.) is wholly superfluous.

At any rate, even if it is granted that Richard's amendment succeeds in rescuing Richard's analysis from the above problem concerning iterated attitude ascriptions, Richard's amendment is *ad hoc*: The essence of the problem is that attitude verbs do not seem to be indexical: when Otho sincerely utters 'I believe that snow is white' in a context c, and Oscar sincerely utters this same sentence in context c', then, contrary to Richard's analysis, it follows that Otho and Oscar hold the same attitude toward the same proposition. Moreover, if Richard's proposal can rescue his analysis from the above difficulty, it does so only at an extremely high price: If the characters of propositional attitudes serve as the Russellian elements of annotations, then RAMs can serve as the "vehicles of truth" only if the predicate 'true' is itself analyzed as an indexical; "'is true' . . . is a predicate that determines a different [property] in different . . . contexts," (p. 246). This complication is both undesirable, and unwarranted.

38 How does Richard's theory cope with ascriptions such as 'The Greeks believed that the Earth was flat'? Richard does not consider such "plural ascriptions," but it seems that in order to account for them Richard must either (a) posit something like Representational Systems for groups of people, or (b) claim that the RAM determined by the that-clause of this ascription is to represent RAMs in the RS of most (many?) individual Greeks. Both options, however, are problematic. Concerning (a), it seems extremely unlikely that there is one set of mental representation types such that all of the relevant Greeks employed instances of those types. More importantly, concerning both (a) and (b), in order to know the truth conditions of 'The Greeks believed that the Earth was flat' one does not even have to consider whether the RAM expressed by the that-clause represents a RAM in the appropriate Greek community RAM. In making ascriptions such as 'The Greeks believed that the Earth was flat' we simply do not consider what sentences of mentalese the Greeks might have employed. How could we citizens of the Twentieth Century even consider such mental representations, especially those of us who do not speak ancient Greek?

39 According to Richard's definition of the character of 'believes', CHAR( ) is function from contexts to *intensions*, i.e. functions from ordered pairs of subjects and RAMs to sets of possible worlds. For the sake of simplicity, I suppose that CHAR( ) is function from contexts to *extensions*, i.e. sets of ordered pairs.

40 Richard 1990, p. 142.

41 See Richard 1990, p. 136.

42 In a footnote in Richard 1992, Richard states that "an adequate account of representations will individuate them so finely that direct--as opposed to quantificational--reference to them will turn out to be (practically) impossible," (p. 131). It is not clear what Richard is proposing here. I assume that by 'quantificational reference' Richard means something like, "reference by description." Is Richard suggesting that in order for Otho to grasp the proposition expressed by Oscar's utterance of (3), Otho and Oscar must *refer* (?!) to the same representational types via the same descriptions? Surely this is implausible. But under any other interpretation,

Richard's proposal does not change anything: The problem is that ordinary speakers do not know enough about one another's representational types to compute the value of R( ) for a given context c. That is, if r is the value of R(c), they cannot "refer" to r "directly," "quantificationally," or any other way, because they lack the ability to distinguish r from r' and r" etc. (Note that if a speaker is to know the truth conditions of a belief ascription, he must know which particular relation is denoted by the occurrence of 'believes'. But this requires him to know which particular restriction is in operation in the context. If r is the restriction determined by R( ), the speaker must somehow be able to distinguish r from other restrictions which might also have been determined.)

43 Richard 1990, p. 83.

44 In denying the Principle of Full Articulation Crimmins and Perry are not committing themselves to the view that there are propositional constituents which are not in any way represented syntactically. That is, Crimmins and Perry allow for the possibility of there being syntactic items which are not phonetically or orthographically realized--something like traces or PRO--which have the relevant propositional constituents as semantic values. They remain neutral as to whether or not there are wholly unrepresented propositional constituents, and are concerned only to deny that all propositional constituents must be overtly represented.

45 Crimmins and Perry maintain that, though notions and ideas are particulars, the same notion or idea can be a constituent in many beliefs (desires, etc.) just as "one may ... be a member of many different committees or clubs," (Crimmins and Perry 1989, ftnt. p. 692.)

46 This is an accurate statement of Crimmins and Perry's account of *explicit*, or *occurrent* belief, though Crimmins offers a different analysis of tacit belief. Crimmins maintains that a subject tacitly holds the attitude of belief toward a Russellian proposition p just in case, roughly, "it is as if A has an explicit belief in p," (Crimmins 1992a, p. 65). To simplify exposition, I will concentrate on Crimmins and Perry's account of explicit belief. (As will become apparent in Chapter 4, I deny that there is any important distinction between tacit and explicit, or occurrent, beliefs; I deny that there are two "kinds" of belief.)

47 Crimmins 1992a, p. 78. Note that, as I understand him, in appealing to the phenomenon of recognition failure Crimmins, unlike Richard, does not violate the Independence Constraint. Richard violates the Independence Constraint because he appeals to the phenomenon of recognition failure as a metaphysical criterion for the identity of representational types. Crimmins, on the other hand, merely states that however notions are in the end to be individuated, they must be individuated at least finely enough to account for the phenomenon of recognition failure. (Because cognitive particulars are physical entities they presumably will be individuated by appeal to physical properties such as causal history and location.)

48 Crimmins and Perry, like Richard, never consider the relation between a subject's verbal behavior and the mental representations the subject instantiates. Thus I am to some extent constructing what I think Crimmins and Perry's responses to the epistemological and semantic arguments would be.

49 This statement is perhaps a bit too strong, as it does not *follow from* the fact that b and b' are tokens that they are not identical, especially since Crimmins and Perry maintain that a notion can appear in more than one belief. (See footnote 45.) It is conceivable that one mental token is associated with both Odile's assent to (O1), and her dissent from (O2). Consequently, even if the mediators of propositional attitudes are tokens, it is not obvious that the individuation and independence constraints will be satisfied. Nonetheless, for the sake of argument I will grant that there is some independent means of individuating cognitive particulars which individuates them finely enough to preclude all nomologically possible epistemological arguments from opacity.

50 Crimmins 1992b, p. 52.

51 This is, unfortunately, only a characterization of Crimmins and Perry's analysis. Much of what I omit concerns how my rather enigmatic phrase 'appropriately composed of' is to be understood. Showing how the content of a belief is determined by its structure and the contents of its constituents is a rather complicated project which extends beyond my somewhat limited concerns here. Also, additional complexities are required in order to account for the truth of ascriptions concerning *tacit* beliefs: On Crimmins' view, it would be true to say of an ordinary subject that he believes that stationwagons are inedible, though it is unlikely that he instantiates a *belief* that has this Russellian proposition as content.

52 Crimmins and Perry 1989, p. 699.

53 For more alleged counterexamples to the Principle of Full Articulation, see Crimmins 1992a, p. 16-21, and Perry 1986. (The arguments against Crimmins and Perry's analysis of attitude ascriptions presented in section 3.3.2 can be generalized to show that none of these phenenoma can adequately be accounted for by appeal to tacit reference and unarticulated constituents. I will not, however, develop this point further.)

54 I am not maintaining that Crimmins and Perry cannot in some way extend there theory so that it could potentially preserve the "literal truth" of such folk explanations of behavior. That is, I am not maintaining that the Deductive-Nomological model of explanation is the *only* model of acceptable explanations of behavior. All I am claiming is that, first, this familiar way of accounting for the truth conditions of explanations is not open to them, and second, if they are to succeed in preserving the "literal truth" of such folk explanations, then the onus is on them to provide an alternative model of explanation.

[55] This is not to say that a more specific location is never intended by an utterance of 'It's raining.' For instance, if I utter this sentence in response to your query, "What's the weather like directly outside your window?" clearly my utterance is true just in case proposition (a) is realized. (Though of course what counts as being "directly outside my window" probably varies from context to context).

Also note that in situations in which no specific location is presented by an occurrence of 'It's raining', our intuitions concerning the truth conditions of the occurrence are quite hazy. For example, if I utter 'It's raining' in an offhand way, and it is raining in the greater Boston area, my utterance seems "truer" than it would be if it were raining directly outside my window only. A correct account of the semantics of 'Its raining' should account for not only the context sensitivity of this sentence, but also the nebulousness of our intuitions concerning the truth conditions of its occurrences. It does not seem, however, that Crimmins and Perry's theory can account for the haziness of our intuitions. If some specific, well defined, place is tacitly referred to by an occurrence of 'its raining', then one would expect ordinary speakers to have firm and well defined intuitions concerning truth conditions.

[56] Note that I am not claiming that there are no such things as *notions*, nor am I claiming that all speakers are too "epistemologically distant" to refer to them. It seems at least plausible to suppose that there are "cognitive particulars" such as *beliefs* and that these entities are composed of entities like notions and ideas. And it even seems plausible that we can, when we have the requisite knowledge and intentions, refer to them. For example, it seems plausible to suppose that the occurrence of 'it' in my utterance of 'It is Quine's normal notion of Clinton' refers to Quine's normal notion of Clinton (if Quine has such a notion). All I am claiming is that it is extremely implausible to interpret *ordinary* speakers uttering *ordinary* attitude ascriptions as *tacitly referring to specific notions and ideas*. The only thing that could determine that I had tacitly referred to *this* particular notion as opposed to *that* particular notion would be my intention to refer to this notion as opposed to that notion, but in ordinary cases speakers lack such intentions.

[57] Perry seems willing to admit that both ascriptions, and negated ascriptions, might express constraints on notions rather than tacitly refer to specific notions. Crimmins, however, maintains that all ascriptions at least attempt to specify notions, though some negated ascriptions merely express constraints on notions. (See note 60.) Crimmins eschews the constraint analysis with regard to (positive) ascriptions because he thinks it is implausible to suppose that specific constraints are provided by occurrences (See Crimmins 1992a, p. 168-9). I, of course, agree with Crimmins on this point, but why does he find it any more plausible to suppose that specific notions, as opposed to constraints, are specified?

[58] Schiffer makes a point quite similar to this in Schiffer 1992.

[59] Crimmins 1992a, p. 183-4.

60 Of course similar arguments would apply if I uttered

(36n) That yokel does *not* believe that it's the tallest thing in the world.

More specifically, we intuit that occurrences of (36n) would be false, rather than meaningless in situation 1, and true rather than meaningless in situation 2.

I find Crimmins' treatment of these difficulties very puzzling. He assumes, without argument, that in either situation 1, or in situation 2, an occurrence of (36) would be meaningless. And then he goes on to amend his account so that in either situation 1, or in situation 2, an occurrence of (36n) is predicted to be true. (He suggests that the 'not' appearing an a relevant occurrence of (36n) designates a "wide-scope"--*it is not the case that*--negation operator.) I do not understand what is motivating Crimmins here: Why is Crimmins willing to abandon our intuitions concerning occurrences of (36), but not (36n)?

61 Crimmins and Perry 1989, p. 705.

62 Crimmins 1992a, p. 44.

4.0 Sketch of a Russellian Theory.

In this chapter I present what I think are the appropriate Russellian responses to the problems and arguments posed by the phenomenon of opacity. That is, I present solutions and responses to the problems and arguments posed by opacity which respect the principles of direct reference and semantic innocence, and also heed the lessons learned in the previous chapters. While these responses do not in and of themselves constitute a full-fledged *theory* of propositional attitudes and propositional attitude ascriptions, they do constitute a kind of a sketch of a viable Russellian theory. The chapter proceeds as follows: In section 4.1 I present and motivate a response to the *epistemological* arguments from opacity, and in section 4.2 I discuss the problem of the cognitive significance of occurrences. In section 4.3 I sketch the theoretical machinery of "Discourse Representation Theory" (or "DRT") and utilize this machinery in presenting a response to the *semantic* arguments from opacity. Finally, in section 4.4 I utilize the machinery of DRT to sketch a Russellian account of the truth conditions of ordinary "folk" explanations of behavior.

4.1 The Appropriate Russellian Response to the Epistemological Arguments From Opacity.

Recall the epistemological arguments from opacity against the Naive Russellian Theory: It follows from Odile's assent to (O1) and

> The Epistemological Principle of Assent: If, at time t, a normal, sincere, understanding subject assents to an occurrence of a declarative sentence Σ, then he holds the attitude of belief toward the content expressed by that occurrence of Σ, at t.

that

(A) Believes <Odile, P>.

(where 'P' designates the Russellian proposition that Twain is a great author.) And it follows from Odile's dissent from (O2) and

> The Epistemological Principle of Dissent: If, at time t, a normal, sincere, understanding subject dissents from an occurrence of a declarative sentence Σ, then he does not hold the attitude of belief toward the content expressed by that occurrence of Σ, at t. (Again note that *dissenting from* an occurrence of a sentence is to be distinguished from *assenting to an occurrence of the negation of a sentence.*)

that

(B) ¬(Believes <Odile, P>).

(I am again ignoring time and tense.) In order to avoid refutation by way of the epistemological arguments, a Russellian theorist must reject either, (i) at least one of the tenets of the Naive Russellian Theory, or (ii) one of the above epistemological principles.

All of the Russellian theorists considered in Chapter 3 opt for (i); they all reject the Naive Russellian Theory's binary analysis of propositional attitudes in favor of a ternary analysis. Thus they avoid refutation by the epistemological arguments by claiming that what follows from Odile's assent to (O1) is not that (A), but rather that

(A*) BEL <Odile, P, m>

(where 'm' designates some kind of mediator--a "mode of apprehension," "representational type," or "cognitive particular.") And they claim that what follows from Odile's dissent from (O2) is not that (B) but rather that

$$(B^*) \quad \neg(BEL <Odile, P, m'>)$$

where 'm' and 'm'' allegedly designate distinct mediators. In this way the Russellian theorists considered in Chapter 3 attempt to avoid the contradictory result that a binary relation both holds and does not hold between Odile and the Russellian proposition that Twain is a great author.

There are, however, several difficulties with this general strategy for avoiding refutation by way of the epistemological arguments. First, as was illustrated in the preceding chapters, it is very difficult to state an acceptable principle of individuation for the posited mediators. (This difficulty is further elaborated in Appendix A.) It seems that the Russellian theorists who posit such mediators are guaranteed to satisfy the individuation and independence constraints only if the mediators are identified with *particulars*. But if the mediators are identified with particulars, as opposed to multiply realizable universals, then the general strategy of positing mediators is rendered *ad hoc*: If mediators are particulars, then there can be no nomonological relationship between mediators themselves and kinds of behavior. Second, the ternary analysis of propositional attitudes does not accord well with the way ordinary subjects speak, and think, about propositional attitudes. This is illustrated by the intuitive validity of arguments such as the following:

> (1) Oscar believes everything that Odile believes.
> (2) Odile believes that Twain is a great author.
> Therefore,
> (3) Oscar believes that Twain is a great author.

It seems natural to analyze this argument as

(1') $\forall x$ ((believes <Odile, x>) $\rightarrow$ (believes <Oscar, x>))
(2') believes <Odile, P>
Therefore,
(3') believes <Oscar, P>

where 'believes' designates the attitude of belief, and 'P' designates the content expressed by 'that Twain is a great author'. Theorists who endorse the ternary analysis of propositional attitudes, however, cannot avail themselves to this analysis; they must either deny that the argument has the logical form displayed in (1')-(3'), or they must deny that occurrences of 'believes' designate the attitude of belief. (Crimmins and Perry seem to opt for the former, while Salmon and Richard opt for the latter.) Both options, however, are counterintuitive: our intuitions dictate that occurrences of propositional attitude verbs designate binary relations and that what they designate are propositional attitudes, not semantic proxies for them.

For these reasons I suggest that it is a mistake to analyze propositional attitudes as ternary relations; the Naive Russellian theory is correct in maintaining that propositional attitudes are binary relations, and that occurrences of attitude verbs designates these binary relations. This is not to deny, however, that propositional attitudes are in some sense "mediated relations" (What would a *nonmediated* relation be?). One must not confuse the question, "What are the relata of propositional attitudes?" with question, "What is it for a propositional attitude to hold between such relata?" Consider, for example, the relation *being North of*. This is a relation which holds between Earth-bound objects, and there is a sense in which it is a mediated relation: x is North of y only if there is some appropriate mediating region of the Earth's surface which lies between x and y. But it would be a mistake to conclude that *being North of* is a ternary relation

which holds between triples consisting of two Earth-bound objects and a mediating region of the Earth's surface; to draw this conclusion would be to confuse issues concerning what the relata of *being North of* are with issues concerning what it is for one of these relata to be *North of* another. To maintain that propositional attitudes must be ternary relations because they are mediated relations would be to make a similar mistake. While it is no doubt true that Odile can hold the attitude of belief toward the Russellian proposition that Twain is a great author only if certain mediating psychological conditions are met, it is a mistake to conclude that attitude of belief must be a ternary relation between objects, propositions, and some aspect of these psychological conditions.

The response to the arguments from opacity which I advocate does not require one to deny the straightforward binary analysis of propositional attitudes and propositional attitude verbs, nor any other tenet of the Naive Russellian Theory. Rather, the response which I advocate is to deny the *Epistemological Principle of Dissent*. In terms of the above argument, I maintain that (B) does not follow from Odile's dissent from (O2). The Epistemological Principle of Dissent, and also the Epistemological Principle of Assent, however, seem to be supported by the intuitions that a subject cannot fail to know the contents of his own mental states. (Hereafter I will refer to these sorts of intuitions as *Cartesian intuitions*.) Though it is difficult to describe these intuitions in detail, the general idea is well expressed by Stalnaker:

> What we see and know is partly a matter of what we are looking at, and what is true, and we can get it wrong. But we can't be wrong about what we think, or think we think about. When I retreat from saying how things are to saying how they seem--how they are *according to me*--I retreat from a claim about the world to a claim about my own mind, and I can tell that the claim is true by introspection--by observing what is internal to my mind.[1]

In view of these intuitions, how can either the Epistemological Principle of Dissent or the Epistemological Principle of Assent be denied? If a subject *understands* an occurrence of Σ and sincerely indicates that he does not believe what is expressed by this occurrence, then how could he possibly be wrong? When a normal understanding subject assents to an occurrence of a sentence (He nods his head, or utters 'true'.) he is indicating that he thinks what is expressed by the sentence is true, and, though the sentence may not in fact express a truth, since he is retreating to a claim about his own mental states, surely he cannot be mistaken. Similarly when a subject dissents from an occurrence of a sentence (He shakes his head, shrugs his shoulders, or utters 'I don't believe it'.), he may be wrong in not thinking that what the sentence expresses is true, but insofar as he is reporting his own mental states, insofar as he is reporting that he does not believe what is expressed by the sentence, surely he must be correct. How can Russellian theorists reject the Epistemological Principle of Dissent without abandoning the Cartesian intuition that a subject cannot be mistaken concerning the contents of his own thoughts?

I think the line of reasoning expressed in the above paragraph is fallacious. Though I heartily endorse our Cartesian intuitions--we are in some sense guaranteed to know the contents of our own thoughts in a direct, nonempirical manner--these intuitions do not provide support for the Principles of Assent and Dissent. In order to understand why this is so, our Cartesian intuitions must be described more precisely: What exactly is it we are guaranteed to know, and in what sense are we guaranteed to know it?

Since the advent of "anti-individualist" or "wide" theories of psychological content, philosophers have been concerned with reconciling our Cartesian intuitions with the fact that the content of one's thoughts may be determined by environmental factors which are beyond one's ken. My belief that *that's a beech tree* is true just in case the object of my belief is a beech tree; if it were an elm tree, my belief would be false, regardless of whether or not I have even heard of elm trees. Given this kind of anti-individualism, how can I be said to know what the content of my belief is? If I cannot

distinguish beeches from elms, how can I know the difference between the thought that *that's a beech tree* and the thought that *that's an elm tree*? And if cannot distinguish the contents of these thoughts, then how can I know that I believe that that's a *beech* tree, as opposed to believing that that's an *elm* tree?

Burge, and others,[2] have provided a solution to this difficulty. Burge maintains that the certainty of our knowledge of the contents of our own thoughts is explained by "cogito-like" judgments, where cogito-like judgments are judgments about the content of one's current thought. Consider the thought that

(a) Water is wet.

To make a cogito-like judgment about one's thought that (a) is to make a self-ascriptive, second order judgment about this thought, viz. the judgment that

(b) I am now thinking that water is wet.

When one makes a cogito-like judgment such as (b), the content expressed by the that-clause is both thought, and thought about, simultaneously. Burge explains,

> When one [judges] that one is thinking that p, one is not taking one's
> thought (or thinking) that p merely as an object. One is thinking that p in
> the very event of [judging] knowledgeably that one is thinking it. It is
> thought and thought about in the same mental act.[3]

Thus, when one makes the cogito-like judgment that

(b) I am now thinking that water is wet.

one simultaneously thinks the thought that

(a) Water is wet.

It is this self-referential feature of cogito-like judgments which accounts for the certainty

of our self-knowledge. Burge explains that "the object, or subject matter, of one's

thoughts is not contingently related to the thoughts one thinks about it. . . . An error

based on a gap between one's thoughts and the subject matter is simply not possible in

these cases."[4] For example, when Bert makes the cogito-like judgment (b), his judgment

must be true because in judging that he is thinking that water is wet, Bert *ipso facto*

thinks that water is wet. Suppose that Bert lived on Twin-Earth where twater--a

substance which has the molecular structure XYZ, yet is phenomenologically similar to

water--ran in the rivers and boiled in the teapots. If Bert inhabited this environment, then

when making the cogito-like judgment that he would express with the sentence 'I am now

thinking that water is wet' Bert would be making the cogito-like judgment that

(c) I am now thinking that twater is wet.

And if Bert's environment were such that (c) is the cogito-like judgment he would be

making, then Bert would ipso facto be thinking a different thought, viz. the thought that

(d) Twater is wet.

In making a cogito-like judgment the content of the that-clause is both thought and

thought about simultaneously, and consequently it is impossible for such cogito-like

judgments to be false; the content of a cogito-like judgment "is self-referentially fixed by

the judgment itself; and the judgment is self-verifying."[5]

It should be clear from the above how Burge's model of self-knowledge manages to reconcile anti-individualism with our Cartesian intuitions. Our knowledge of the content of our own thoughts is compatible with anti-individualistic theories of content because no facts about the environment of a subject can impugn the correctness of a cogito-like judgment made by that subject. In order to think an empirical thought certain physical facts about one's environment, or *enabling conditions*, must obtain. For example, in order for Bert to think a thought with the content *twater is wet*, it must be XYZ that runs in the rivers and boils in the teapots of Bert's environment. One does not, however, need to know the complex environmental facts which determine the content of one's thought in order to think an empirical thought. In order to think that twater is wet Bert need not have the capacity to distinguish XYZ from H_2O, nor must he know that the content of his thought is individuated by his relation to certain physical and/or social facts. One need not be a chemist and a philosopher to think that water is wet.

The same considerations hold for cogito-like judgments. If Bert makes a cogito-like judgment, then the enabling conditions for making that judgment are met, regardless of what these conditions involve. Bert need not have the capacity to distinguish H_2O from XYZ in order to make the cogito-like judgment that

> (b) I am now thinking that water is wet.

And if (b) is the cogito-like judgment that Bert would make in his environment, then in making this judgment Bert correctly identifies the content of his thinking. The necessary correctness of cogito-like judgments is not impugned by anti-individualism because, Burge explains,

> If background conditions are different enough so that there is another
> object of reference in one's self-referential thinking, they are also different
> enough so that there is another thought.[6]

Burge's account of self-knowledge goes a long way toward explaining and clarifying our Cartesian intuitions. It explains why we are guaranteed to know, in a non-empirical manner, the content of our thoughts and it does so without positing suspicious mental entities--Cartesian ideas, Empiricist sense data, or Fregean senses--that are in some way "epistemologically transparent." Furthermore, Burge's account successfully explains how we are guaranteed to know the content of our thoughts in this way even though the content of our thoughts may be fixed by factors beyond our ken. Notice, however, that it does not follow from Burge's account that a subject is guaranteed to know whether or not he holds a particular *propositional attitude* toward the content of his thinking. Thinking, or better *entertaining*, a particular content is to be distinguished from *holding the attitude of belief*, or any other attitude, toward the content of one's thinking. When a subject is thinking that p in the very act of judging knowledgeably that he is thinking it, he cannot misidentify the content of his thoughts, but it does not follow that he cannot be mistaken concerning the attitude he holds toward that content. As Davidson states, "the argument shows only that one cannot go wrong in identifying the content of an attitude, which is not a reason why one cannot go wrong about the existence of the attitude."[7] Burge's account of self-knowledge explains why a subject cannot fail to know the contents of his own thoughts, yet it does not imply that a subject is guaranteed to know whether or not he holds a particular attitude toward a particular content. Consequently, assuming that Burge's account of self-knowledge is roughly correct, the phenomenon of guaranteed self-knowledge does not provide support for the epistemological principles of assent and dissent. Thus Russellian theorists can reject the Principle of Dissent without abandoning our Cartesian intuitions.

Though neither the Epistemological Principle of Assent nor the Epistemological Principle of Dissent is supported by the phenomenon of guaranteed self-knowledge, there are cogent reasons for endorsing the Epistemological Principle of Assent. I will argue

that there is a necessary connection between a subject's sincere understanding *assent to* an occurrence of Σ, and the subject's holding the attitude of belief toward the content expressed by this occurrences of Σ, and thus the Epistemological Principle of Assent enjoys a kind of privileged status. There is, however, no such necessary connection between a subject's sincere understanding *dissent from* an occurrence of Σ and the subject's *not* holding the attitude of belief toward the content expressed by this occurrence of Σ.

Recall the Epistemological Principle of Assent:

> The Epistemological Principle of Assent: If, at time t, a normal, sincere, subject who understands an occurrence of Σ assents to this occurrence, then he holds the propositional attitude of belief towards the content expressed by this occurrence of Σ, at t.

Why is it that if a sincere understanding subject assents to an occurrence of Σ, then he *must* hold the attitude of belief toward the content expressed by this occurrence? What is the necessary connection between sincere understanding assent to an occurrence, and holding the attitude of belief toward the content expressed by the occurrence? It is significant that an act of assenting to an occurrence need not involve a self-ascription; in assenting to an occurrence of Σ, one need not utter, nor think a thought that would be expressed by, a sentence of the form ⌜I believe that Σ⌝ Rather in assenting to an occurrence one need only nod and judge that the content expressed by Σ is true. Thus the Epistemological Principle of Assent is true not because of a ubiquitous ability of subjects to know, and report, (i) the contents of their propositional attitudes and (ii) which propositional attitudes they hold toward those contents, but rather the principle is true because one's act of sincere assent to an occurrence of Σ constitutes conclusive evidence that one holds the attitude of belief toward the content expressed by this occurrence. In order to motivate and clarify this defense of the Epistemological Principle of Assent, I

must at least adumbrate an account of what it is for a subject to hold a propositional

attitude--the attitude of belief in particular--toward a Russellian proposition.

Though I cannot give a comprehensive analysis of propositional attitudes, I

endorse a dispositional and broadly functional analysis. More specifically, I agree with

Stalnaker's rough characterizations:

> To desire that P is to be disposed to act in ways that would tend to bring it
> about that P in a world in which one's beliefs, whatever they are, were
> true. To believe that P is to be disposed to act in ways that would tend to
> satisfy one's desires, whatever they are, in a world in which P (together
> with one's other beliefs) were true.[8]

(I optimistically maintain that all other propositional attitudes, e.g. hope, fear, regret,

etc., can be defined in terms of belief and desire. Also note that if Odile satisfies these

conditions for belief with regard to the Russellian proposition presented by (O1), then

she also satisfies these conditions with regard to the Russellian proposition presented by

(O2): If Odile is disposed to do what would best satisfy her desires if the proposition

presented by (O1) were true, then Odile is *ipso facto* disposed to do what would best

satisfy her desires if the proposition presented by (O2) were true.)

Hence I maintain that a subject holds the attitude of belief toward P if and only if

the subject is disposed to behave in certain ways. To maintain this dispositional analysis

of the attitude of belief, however, is not to deny that some kind of intrinsic

psychological\neurological properties must be instantiated by a subject if he is to hold the

attitude of belief toward a Russellian proposition. Just as a physical object has the

dispositional property of being brittle only if it instantiates certain intrinsic

physical\chemical properties, so a subject holds the attitude of belief toward a Russellian

proposition only if he instantiates certain intrinsic neurological\psychological properties.

(Moreover, just as an ordinary subject can know that an object is brittle even though he

has no detailed knowledge of the intrinsic properties responsible for its brittleness, so an

ordinary person can know of a subject that he holds the attitude of belief toward P even though the ordinary person has no detailed knowledge of the intrinsic properties responsible for this belief.) Furthermore, the intrinsic properties responsible for a subject's holding the attitude of belief toward a Russellian proposition may, and I assume do, involve the instantiation of systems of mental representations which represent objects and properties. (Consequently this dispositional analysis of propositional attitudes is wholly compatible with a "representational theory of the mind.") This is not to suggest, however, that holding the attitude of belief toward P requires that a subject "token" a "sentence of mentalese" which represents P. As Stalnaker suggests, "these representations could conceivably take the form of sentences of a language of thought written in the belief center in one's brain, but they could also take the form of pictures, maps, charts, or graphs, or (most plausibly) a diversity of redundant forms, none of which are very much like any of the forms which our public representations take."[9]

Thus to hold the attitude of belief toward P is to be disposed to behave in such a way that, if one's beliefs are (or were) true, one's desires will be (or would be) satisfied. And I assume that one can be so disposed only if one instantiates an appropriate mental representation, or a system of mental representations, i.e. a mental representation whose content is P. But what is it for a mental representation, or more generally a state of a representational system, to have Russellian proposition P as its content? Not surprisingly, I can give only the barest sketch of an answer to this question. Following Dretske (more or less[10]), I maintain that an instance of a state σ of a representational system (where humans are but one kind of representational system) *represents*, or has as content, Russellian proposition P just in case instances of σ are the result of a process or mechanism whose *function* is to *indicate* P by producing or resulting in instances of state σ. (*Misrepresentation* of P occurs when the representational system instantiates an instance of a state σ which is produced by a mechanism whose function is to indicate P by producing instances of σ, though P does not in fact obtain.) A process or mechanism

indicates P by producing or resulting in instances of a state σ just in case it is

nomologically necessary that the mechanism produce an instance of σ if and only if P

obtains. For a process or mechanism to have as its *function* indicating p by producing

instances of σ is for the process or mechanism to be, in some sense, designed to indicate

P by producing instances of σ--indicating P by producing σ is what the process or

mechanism is *supposed to do*. For example, a fuel gauge has the *function* of indicating

that the tank is half full by producing instances of states in which the needle points to the

'1/2' symbol. (Note that, if the gauge is working properly, instances of this state may

indicate not only that the tank is half full, but also that the battery has at least a certain

charge, and that the tank is placing a certain gravitational force on the bolts which secure

it, etc.. Though an instance of this state in a working fuel gauge may *indicate* all of these

propositions, it *represents* only the proposition that the tank is half full because the fuel

gauge is not designed to indicate these other propositions.) Similarly, I boldly assume,

an instance of a mental state σ' instantiated by a subject who has just looked at the cat

lying on the mat represents that the cat is on the mat because this instance of σ' is

produced by a (very complex) process or mechanism whose function is to indicate that

the cat is on the mat. (Of course instruments like fuel gauges have the functions they do

because we assign them those functions, and thus the functions of such instruments are

derivative of the functions of the processes which produce or result in our mental states.

But what determines the *functions* of the processes which produce our mental states? *If*

this question has an answer, I suspect that it has something to do with how our cognitive

faculties which produce these states develop through learning and natural selection: A

cognitive mechanism may have the function of indicating P by producing instances of σ

because the mechanism's existence in an organism can be explained by the fact that,

under "ordinary circumstances," the mechanism produces instances of σ if and only if P

obtains.[11] Let us leave it at that.)

How and why does the Epistemological Principle of Assent follow from this, admittedly very sketchy, analysis of propositional attitudes and their contents? Why does it follow--*ceteris paribus*--from a normal subject's sincere understanding assent to an occurrence of Σ that he holds the attitude of belief toward the content expressed by Σ? The answer is that in engaging in the act of sincere understanding assent to an occurrences of Σ, the subject manifests a disposition which is sufficient for holding the attitude of belief toward the content expressed by Σ. That is, in engaging in the act of sincere understanding assent to an occurrences of Σ, the subject behaves in such a way that, if his beliefs are true, his desires will tend to be satisfied. For example, Odile may desire to convince Oscar that she knows something about American literature, and she may believe that Twain is a great author and that this is a fact concerning American literature. If these beliefs are true, then Odile's assent to (O1) in Oscar's presence would, under normal circumstances, satisfy her desire. Consequently, an act of sincere understanding assent to an occurrence of Σ which presents P is one way in which a disposition sufficient for holding the attitude of belief toward P may be manifested. (Moreover, since by assumption the assenting subject *understands* the occurrence of Σ, it is plausible to assume that the subject has this disposition because he instantiates an internal mental state whose content is the content expressed by the occurrence of Σ; understanding an occurrence of Σ which expresses content P would seem to require instantiating a state which represents P.) Thus in assenting to an occurrence of Σ one does not correctly *report* that one believes the content expressed by Σ; rather the very act of sincere understanding assent is a manifestation of one's belief in the content expressed by Σ.[12]

The Epistemological Principle of Dissent does not, however, enjoy this privileged status, but before I present my arguments against the Epistemological Principle of Dissent, a few clarifying remarks concerning the Epistemological Principle of Assent are in order. First, note that both the "sincere," and the "understanding" provisos in the

antecedent of the principle are required: An *insincere* assent to an occurrence is not a manifestation of the appropriate disposition, and neither is an assent made in ignorance of the content expressed by the occurrence. Second, note that while Russellian theories cannot be refuted by epistemological arguments from opacity based upon the Epistemological Principle of Assent alone (the Epistemological Principle of Dissent is also required), it does follow from the Epistemological Principle of Assent alone that subjects can, and often do, hold the attitude of belief toward contradictory--and therefore impossible--Russellian propositions. For example, it follows from the Epistemological Principle of Assent and Odile's sincere understanding assent to an occurrence of

(1&2n) Twain is a great author and Clemens is not a great author.

that Odile holds the attitude of belief toward a contradictory, and therefore impossible, Russellian proposition. (For a Russellian proposition to be *impossible* is for it to obtain in no possible world. I assume that Twain is necessarily identical to Twain.) I think that this feature of the Russellian view is intuitively correct: Under ordinary circumstances, if the ancient astronomer behaves as if Phosphorus rises in the morning and does not rise in the morning, then he holds the attitude of belief toward the contradictory and impossible proposition that Phosphorus rises in the morning and Hesperus does not rise in the morning. And if Odile assents to an occurrence of (1&2n), then she holds the attitude of belief toward the contradictory and impossible proposition that Twain is a great author and Clemens is not a great author. Moreover, these contradictory beliefs may be crucial for explaining some of our behavior because often our behavior is, in a way, contradictory. Suppose, for example, that on one occasion Odile sees Twain and asks him for his autograph, while on another occasion she sees Twain and thumbs her nose at him. What other than a contradictory belief (or contradictory beliefs) could account for this strange behavior?

Despite the intuitive plausibility of belief in contradictory and impossible propositions, however, this feature of the Russellian view is often taken to be a reason for rejecting the Russellian view of propositions. First, it is sometimes objected that it cannot be correct to attribute contradictory beliefs to a subject because (normal) subjects are *rational*, and believing in contradictions is *irrational*. I will not here undertake an investigation of what it is to be rational, but this objection clearly begs the question against a Russellian theory: Clearly a Russellian theory maintains either (i) that (normal) subjects are often irrational, and rationality is a kind of goal to which ordinary subjects aspire, or that (ii) a (normal) subject can be rational even if he holds the attitude of belief toward contradictory propositions. The first option would involve appeal to a "wide" analysis of rationality; under the first option rationality would involve consistency among the Russellian propositions believed. The second option, on the other hand, would involve appeal to a kind of "narrow" rationality; rationality would involve some kind of consistency in the mental representations instantiated by a subject.[13]

A more serious objection to the Russellian view concerns *what it is* to hold the attitude of belief (or any other attitude) toward an impossible Russellian proposition. According to the analysis of belief advocated above, to hold the attitude of belief toward the Russellian proposition P is to be disposed to behave in such a way that *if one's beliefs were true*, one's behavior would tend to bring about the satisfaction of one's desires. But if the content of one's belief is impossible, then how are we to understand the above counterfactual? What are we supposing the world to be like when we suppose, for example, that Odile's belief that Twain is not Clemens is true? Since it is necessarily true that Twain is Clemens, what we are supposing cannot be identified with a possible state of the world. One way around this difficulty would be to deny that the Epistemological Principle of Assent holds in cases where the relevant occurrence presents an impossible proposition. Thus a subject's sincere understanding assent to an occurrence of 'there exists a procedure for trisecting a Euclidean angle' would be treated in the same way that

a sincere understanding assent to an occurrence of 'Santa is fat' would be treated. In neither case does it follow from the subject's sincere understanding assent that he holds the attitude of belief toward a Russellian proposition: in the latter case Σ does not present a Russellian proposition at all, and in former case--it might be claimed--Σ presents an inappropriate Russellian proposition. While this is an option, it is not an option I readily adopt. For the reasons discussed above, it seems that (normal) subjects do hold propositional attitudes towards impossible propositions. I leave this is an unsolved problem for the Russellian view.

I have argued that the Epistemological Principle of Assent enjoys a privileged status because there is a special connection between a subject's sincere understanding assent to Σ and the subject's holding the attitude of belief toward the content expressed by Σ: A subject's assent to Σ is, *ceteris paribus*, a sufficient condition for the subject's holding the attitude of belief toward the content expressed by Σ because to assent to Σ is to manifest a disposition which is sufficient for holding the attitude of belief toward the content expressed by Σ. There is, however, no similar reason for endorsing the Epistemological Principle of Dissent. Recall the principle:

> The Epistemological Principle of Dissent: If a normal, sincere, subject
> who understands an occurrence of Σ dissents from this occurrence at t,
> then he does not hold the propositional attitude of belief toward the
> content expressed by this occurrence, at t.

An act of sincere understanding assent to an occurrence of Σ is a manifestation of one's holding the attitude of belief toward the content expressed by Σ. But *dissenting from* an occurrence of Σ does not in the same way constitute a manifestation of one's not holding the attitude of belief toward the content expressed by Σ. If a sincere understanding subject dissents from an occurrence of Σ, and thereby *refrains* from manifesting a disposition which is sufficient for holding the attitude of belief toward the content of

expressed by Σ, it does not follow that he does *not* hold the attitude of belief toward this content. Granted, the subject does not, in the very act of sincere understanding dissent from Σ, *manifest* a disposition sufficient for holding the attitude of belief toward the content of Σ. But of course it does not follow from the fact that the subject does not in the act of dissenting manifest such a disposition that he does not *have* such a disposition. For example, Odile dissents from (O2) and thereby *refrains from manifesting* a disposition sufficient for holding the attitude of belief toward the Russellian proposition that Twain is a great author. Yet in assenting to (O1) and, we may suppose, buying all of Twain's books, Odile does at other times manifest a disposition sufficient for holding the attitude of belief toward the proposition that Twain a great author. Hence Odile's dissent from (O2) is compatible with her holding the attitude of belief toward the content expressed by (O2).

What, if anything, does follow from a subject's sincere understanding dissent from an occurrence of Σ? In some cases a subject might dissent from an occurrence of Σ because he holds the attitude of belief toward the content expressed by the negation of Σ. This does not, however, hold in general. For example, suppose that Odile is just not sure whether Twain is a great author or not. If Odile were in this state of perplexity, she would dissent from an occurrence of

(1) Twain is a great author.

even though she does not hold the attitude of belief toward the Russellian proposition presented by an occurrence of

(1n) Twain is not a great author.

227

Moreover, even it did follow from a subject's sincere understanding dissent from an occurrence of Σ that he holds the attitude of belief toward the content expressed by an occurrence of the negation of Σ, this would not engender epistemological arguments from opacity. It is not contradictory to claim that Odile holds the attitude of belief toward the Russellian proposition that Twain is a great author and also the Russellian proposition that Twain is not a great author. If the attitude of belief is a mediated binary relation between subjects and Russellian propositions, one should expect this kind of conflict in belief contents to be relatively common. (This is not to deny that in most contexts it would be extremely odd to *say* "Odile believes that Twain is a great author and she also believes that Twain is not a great author." But the oddness of such an ascription concerns the semantics and pragmatics of attitude ascriptions, not the nature of propositional attitudes themselves.)

One thing that does, in most cases, follow from a subject's sincere understanding dissent from an occurrence of Σ is that the subject does not hold the attitude of belief toward the Russellian proposition that this very occurrence of Σ is true. For example, in most cases, it does follow from Odile's dissent from (O2) that she does not hold the attitude of belief toward the Russellian proposition that (O2) is true. If a sincere understanding subject believes about an occurrence which he understands that it is true, and he is sincere, then, usually, he will assent to it. (The qualifications are required because it is, I suppose, possible for a subject to dissent sincerely from an occurrence which he understands even though he has a disposition sufficient for holding the attitude of belief toward the proposition that the occurrence is true.) Similarly it follows from a subject's sincerely assenting to an occurrence of Σ that he believes about that occurrence that it is true. Again, if the subject did not believe that the occurrence were true, then, other things being equal, he would not have assented.[14] But, again, no contradiction is engendered by these implications: I maintain that it follows from Odile's assent to (O1) that

(O) Believes <Odile, R>

where 'R' designates the Russellian proposition that (O1) is true. And I maintain that it follows from Odile dissent from (O2) that

(P) ¬(Believes <Odile, S>)

where 'S' designates the Russellian proposition that (O2) is true. Since (O1) is not identical to (O2), R and S are distinct Russellian propositions and thus (O) and (P) are not contradictories. (The claim that sincere understanding acts of assent and dissent require the presence and absence of certain background meta-linguistic beliefs, respectively, is important because these background beliefs can be invoked by Russellian theorists to account for the cognitive significance of occurrences. If an act of sincere understanding assent to an occurrence requires that the subject believe that occurrence is true, then the Russellian theorist can provide ordinary "folk" explanations of why, e.g., Odile assents to (O1), yet dissents from (O2): she assents to (O1) because she believes that (O1) is true, and she dissents from (O2) because she does not believe that (O2) is true. These issues concerning the cognitive significance of occurrences are further explored in the next section.)

I conclude that Russellian theorists can reject the Epistemological Principle of Dissent without denying our Cartesian intuitions, and furthermore, that while there are cogent reasons for endorsing the Epistemological Principle of Assent, there are no corresponding reasons for endorsing the Epistemological Principle of Dissent. What then explains the intuitive appeal of the Epistemological Principle of Dissent? I suggest that at least some of the intuitive appeal of the Principle of Dissent is due to a confusion between the *epistemological* claim the principle actually makes, and a similar sounding

claim concerning what can truly be *said* of a subject who has just dissented from an occurrence of a sentence; i.e. the Epistemological Principle of Dissent is confused with the Semantic Principle of Dissent. We do, in many cases, have the intuition that it would be false to make an ascription of the form $\ulcorner$N believes that $\Sigma\urcorner$ relative to a context in which it is known that the referent of N has just dissented from an occurrence of Σ. But these intuitions concerning the truth conditions of ascriptions do not support the Epistemological Principle of Dissent unless an occurrence of an ascription of the form $\ulcorner$N believes that $\Sigma\urcorner$ is false only if the referent of N does not hold the attitude of belief toward the content expressed by Σ. As will be shown in section 4.3, however, the Russellian theorist is not compelled to grant such an assumption.

4.2 The Appropriate Russellian Response to the Problem of Accounting for the Cognitive Significance of Occurrences.

The problem of accounting for the cognitive significance of occurrences is the problem of explaining a subject's judgments and behavior with regard to occurrences; e.g. "Why does Odile assent to (O1), yet dissent from (O2)?" Or more specifically, "What is it about these occurrences which is relevant to explaining Odile's judgments and behavior with regard to these occurrences?" Providing a *bona fide* solution to this problem would require a comprehensive theory of human judgment and behavior and, of course, I know of no such theory. Consequently my goals in this section will be somewhat modest. Rather than attempting to provide a *bona fide* solution to the problem of cognitive significance, I will merely attempt to show that, contrary to what is often alleged, Russellian theories are no worse off with regard to this problem than are Fregean theories.

A brief review is in order. The Fregean approach maintains that it is the *content* expressed by an occurrence for a subject which accounts for the cognitive significance of

the occurrence for that subject. For example, the Fregean approach maintains that Odile assents to (O1) yet dissents from (O2) *because* these occurrences express distinct *thoughts* for her, and thoughts, which are composed of *cognitive values*, serve as the objects of propositional attitudes. Russellian theories, on the other hand, must deny that the cognitive significance of an occurrence for a subject is accounted for by the content the occurrence expresses for the subject. Russellian theorists maintain that (O1) and (O2) express the same content--the same Russellian proposition--for Odile. Consequently they cannot account for Odile's differing behavior with regard to (O1) and (O2) by appeal to the content(s) expressed by these occurrences; it makes no sense to maintain that Odile assents to (O1) yet dissents from (O2) because of the contents expressed by (O1) and (O2), when these occurrences express the same content for her.

What, according to my view, accounts for the cognitive significance of occurrences? I think Perry is correct in charging that the Fregean account of cognitive significance commits the "fallacy of misplaced information"; viz. "The idea that all the information in an utterance must come from [the proposition presented and/or the content expressed.][15] Our problem is to explain why Odile, a rational agent with a full stock of beliefs, desires and fears, behaves as she does in response to the perception and comprehension of two distinct occurrences, two distinct physical events. Why should we suppose that, in this rich and complex environment, Odile's behavior with regard to (O1) (O2) must be determined solely by whether or not she holds a particular attitude toward the content expressed by these occurrences? Why should the Russellian theorist assume, along with the Fregean theorist, that the fact that Odile "grasps" the contents of (O1) and (O2) by way of perceiving distinct physical objects is irrelevant to Odile's behavior? Surely the fact that Odile is responding to two qualitatively distinct occurrences is relevant to explaining her behavior. Think of Odile as a black box which, given input, produces output. Given the perception (and comprehension) of (O1) as input, assent is the output and given the perception (and comprehension) of (O2) as input, dissent is the

output. Our task is to find the difference in the inputs which explains the differences in output. (Thus we assume that nothing, or nothing relevant, changes inside the box; i.e. Odile does not "change her mind," or become insane, etc., between the time of her perception of (O1) and the time of her perception of (O2).) One suggestion, the Fregean suggestion, is that (O1) and (O2) express distinct contents for Odile, and this difference in expressed content is the relevant difference between the inputs. But we certainly are not constrained to posit such a difference in content in order to account for the difference in output. Moreover, given that, (i) there is already an obvious difference between the inputs, viz. (O1) is numerically and qualitatively distinct from (O2), and (ii) there is good reason to suppose that (O1) and (O2) express the same content, i.e. there are good reasons to endorse the Principle of Direct Reference, it seems that the Fregean account of cognitive significance is neither the simplest, nor the most plausible account available.

I propose that the problem of the cognitive significance of occurrences can be solved by appeal to the acts of perception and comprehension presupposed by instances of the phenomenon of opacity; Odile assents to (O1) yet dissents from (O2) because the event whereby Odile perceives and comprehends (O1) is different in important ways from the event whereby Odile perceives and comprehends (O2). Of course in saying this, I fall far short of actually providing a solution to the problem of cognitive significance; as was mentioned above, providing a *bona fide* solution to this problem would require me to show how my approach can be incorporated in a comprehensive, and true, theory of human psychology, and I know of no such theory. In order to accomplish my goal of showing that the above adumbrated Russellian approach is no less plausible than the Fregean approach, however, I only need to illustrate that it is just as plausible to suppose that the Russellian approach can be incorporated in a comprehensive theory of human psychology as it is to suppose that the Fregean approach can be so incorporated. My strategy for doing this will be as follows: I assume that a comprehensive theory of human judgment and behavior will be either a theory in the

domain of Cognitive Science whose laws are couched in wholly non intentional, terms, or it will be an intentional theory, a theory whose laws take the form of familiar belief/desire explanations. (I also assume physicalism: I assume that either kind of theory will be compatible with a materialist metaphysics.) I will attempt to show that with regard to either kind of comprehensive theory, my Russellian approach is just as likely to be incorporated as is the Fregean approach.[16]

The Fregean approach maintains that Odile assents to (O1) yet dissents from (O2) *because* these occurrences express distinct thoughts for her. How would this general strategy for accounting for the cognitive significance of (O1) and (O2) be realized by a comprehensive non intentional psychological theory? The idea must be something like this: Where T and T' are the thoughts expressed by (O1) and (O2), respectfully, Odile's grasping T requires Odile to instantiate psychological state σ, and Odile's grasping T' requires Odile to instantiate psychological state σ'. (As I argued in Chapter 1, if the Fregean strategy is to be at all plausible from a psychological point of view, it must appeal to psychological states such as σ and σ'; unless T and T' are identified with, or mapped one-to-one with, σ and σ' respectively, the fact that T is distinct from T' is completely irrelevant to the explanation of Odile's behavior. Note that σ and σ' can be *brain-states, functional-states*, or any other kind of state compatible with physicalism.) Odile's assent and dissent is then to be explained by invoking laws something like the following:

A Fregean Law of Assent: For all subjects X, if conditions C are satisfied, and X instantiates σ, then X will assent.

A Fregean Law of Dissent: For all subjects X, if conditions C are satisfied, and X instantiates σ', then X will dissent.

(Conditions C specify--in non intentional, physicalistic, terms--all other relevant features of Odile's psychological state. Thus the same conditions C appear in both the Fregean Law of Assent, and the Fregean Law of Dissent. In terms of the black-box metaphor, that the same conditions are called for in both laws ensures that nothing inside the box changes and thus that the difference in output must be due to a difference in input.) Thus if the Fregean approach is to be realized by a comprehensive non intentional psychological theory, physical states such as σ and σ' which are correlated one-to-one with Fregean thoughts must be identified, and laws such as those stated above must be discovered. Moreover, if the theory is to have any explanatory and predictive power, the psychological states appealed to must be individuated independently of the behaviors they are alleged to explain. The task of discovering and formulating such a theory is, of course, rather daunting.

How would the Russellian approach adumbrated above be incorporated into a comprehensive reduced psychological theory? Just as the real explanatory work on the Fregean approach is done by the states σ and σ' which are required for the "grasping" of Fregean thoughts T and T', so on the Russellian approach the real explanatory work is done by mediating states μ and μ', which are the psychological states Odile instantiates in the perception of (O1) and (O2), respectfully. Mediating states are, roughly, the output of our perceptual modules and the input to our linguistic modules. (Where, following Fodor, a "module" is an "informationally encapsulated computational system"[17] which is responsible for a particular cognitive ability.) Assuming that an ordinary speaker's linguistic knowledge is somehow implemented in a module of his brain, the inputs to this module in part consist of representations of occurrences he has perceived. These representational inputs are tokens which, I assume, can be grouped into types, or *mediating states*.[18] It is important to notice that mediating states need not beanything like "sentences of mentalese": In perceiving an occurrence a subject instantiates an instance of a mediating state, and thus there is a sense in which instances of mediating

states are representations of occurrences--mediating states are not representations of people, cities, properties, etc. (Nor, however, am I committed to the view that there is no such thing as mentalese; as Fodor and other representationalists have argued, some kind of "internal language" may be required in order to explain certain cognitive phenomena.)

How, exactly, might mediating states be invoked by a comprehensive reduced psychological theory in order to account for the cognitive significance of (O1) and (O2) for Odile? If Odile perceives and comprehends (O1), then in the perception and comprehension of (O1) Odile instantiates a mediating state μ. And in perceiving and comprehending (O2), Odile instantiates a distinct mediating state μ'. Odile assents to (O1) yet dissents from (O2) because in perceiving and comprehending these occurrences, Odile instantiates distinct mediating states. More specifically, just as the Fregean approach is committed to there being laws such as those displayed above, so the Russellian approach is committed to there being laws something like the following:

> A Russellian Law of Assent: For all subjects X, if conditions C are satisfied, and X instantiates μ, then X will assent.

> A Russellian Law of Dissent: For all subjects X, if conditions C are satisfied, and X instantiates μ', then X will dissent.

(Again, conditions C specify--in physicalistic, non intentional terms, all other relevant features of Odile's psychological state. In terms of the black-box metaphor, that conditions C are invoked in both laws ensures that the only relevant change between Odile's assent and her dissent concerns μ and μ', the inputs.) The Russellian theorist who opts for the approach sketched above thus faces a daunting task which is in all relevant ways similar to the daunting task facing the Fregean theorist: the Russellian theorist must discover a comprehensive psychological theory which appeals to mediating states such as μ and μ', and, if the theory is to have any explanatory and predictive power, these

states must be individuated independently of the behaviors they are alleged to cause. Thus both the Fregean and the Russellian face similar daunting tasks if they are to provide a solution to the problem of the cognitive significance, and therefore there is no reason, at this point, to prefer one approach over the other. (Of course, as psychology and cognitive science progress, one or the other approach to the problem of cognitive significance may begin to look more plausible, and I hinted above that the Russellian approach may have the edge due to its simplicity. But the question of which approach, if either, will prevail in the end is an empirical one, and thus cannot be answered by armchair ridden philosophers.[19]) I conclude that if the Fregean approach is better suited to account for the cognitive significance of occurrences, it is not because it is more amenable to a non intentional comprehensive psychological theory.

Perhaps then, the Fregean approach to the problem of cognitive significance is superior to Russellian approach adumbrated above not because it is more amenable to a *non intentional* comprehensive psychological theory, but rather because the Fregean account of cognitive significance is more amenable to a comprehensive *intentional* psychological theory--a theory which utilizes familiar belief/desire explanations of behavior. After all, the Fregean approach contends that Odile assents to (O1) *because* this occurrence expresses the content T, and Odile holds the attitude of belief toward this content. So it would seem that the Fregean approach might be more amenable to familiar intentional explanations of behavior than is the Russellian approach. Contrary to initial appearances, however, the Fregean approach is not superior to the Russellian approach in this regard either.

Intentional belief\desire explanations of behavior instantiate the following schema:

(i) N DESIRED that Σ.
(ii) N BELIEVED that if he did A, his DESIRE that Σ would be satisfied.
So,
(iii) N did A.

Attitude verbs are capitalized in the above schema to remind the reader that we are not here doing semantics; we are not here trying to preserve the intuitive truth of certain "folk" explanations of behavior. Rather in this context we are attempting to provide scientific explanations of Odile's behavior. That is, we are trying to determine whether the Russellian theorist can state belief\desire explanations which are true *under his own analysis*. Consequently our untutored intuitions concerning the truth of these explanations of behavior are in this context irrelevant. Thus the above schema and instances of it are stated in theory-neutral terms: Under the Russellian interpretation, 'DESIRED' designates a relation which holds between subjects and Russellian propositions, and that Σ designate the Russellian proposition expressed by Σ. Under the Fregean interpretation, on the other hand, 'DESIRED' designates a binary relation between subjects and thoughts, and that Σ designates a Fregean thought allegedly expressed by Σ. I will argue that, first, there are belief\desire explanations of Odile's behavior--instances of tl ·bove schema--which are true under the Russellian interpretation. And second, the belief\desire explanation of Odile's behavior which would be given by the Fregean theorist is true under the Russellian interpretation. Thus not only can the Russellian provide belief\desire explanations of Odile's behavior, but furthermore the Russellian can endorse the instances of the above schema which would be endorsed by the Fregean approach.

The Russellian approach adumbrated above can utilize the following instance of the schema in order to explain Odile's behavior with regard to (O1) and (O2):

Intentional Explanation A

Part 1.
(i) Odile DESIRED that she indicate that she BELIEVED that (O1)
conveyed a truth.
(ii) Odile BELIEVED that if she assented to (O1), she would succeed in
indicating that she BELIEVED that (O1) conveyed a truth.
So,
(iii) Odile assented to (O1).

Part 2.
(i) Odile DESIRED that she indicate that she did not BELIEVE that (O2)
conveyed a truth.
(ii) Odile BELIEVED that if she dissented from (O2), she would succeed
in indicating that she did not BELIEVE that (O2) conveyed a truth.
So,
(iii) Odile dissented from (O2).

The Russellian can endorse Intentional Explanation A because lines (i) and (ii) for both

part 1 and part 2 are true under the Russellian interpretation. Hence, if we assume the

legitimacy of instances of the belief\desire explanation schema, under the Russellian

interpretation both parts of explanation A provide legitimate explanations. Therefore the

claim that the Russellian approach to cognitive significance is not amenable to intentional

belief\desire explanations is simply false.

Of course Intentional Explanation A is true under either the Russellian or the

Fregean interpretation, and it is true under the Russellian interpretation because the very

occurrences (O1) and (O2) appear in the contents of the attitudes ascribed to Odile. Thus

the Fregean theorist might claim that, while the Russellian can give some belief\desire

explanations, he cannot endorse *as many* as the Fregean theorist. That is, it might be

thought that Fregean approach to cognitive significance is superior to the Russellian

approach because the Fregean theorist, but not the Russellian theorist, can endorse the

following belief\desire explanation:

<u>Intentional Explanation B</u>

Part 1.
(i) Odile DESIRED that she indicate that she BELIEVED that Twain is a great author.
(ii) Odile BELIEVED that if she assented to (O1), she would succeed in indicating that she BELIEVED that Twain is a great author.
So,
(iii) Odile assented to (O1).

Part 2.
(i) Odile DESIRED that she indicate that she did not BELIEVE that Clemens is a great author.
(ii) Odile BELIEVED that if she dissented from (O2), she would succeed in indicating that she did not BELIEVE that Clemens is a great author.
So,
(iii) Odile dissented from (O2).

I maintain, however, that the above claim is false. That is, I maintain that if the above explanation is true under a Fregean interpretation, then it also true under a Russellian interpretation. (I also maintain that Explanation A is a better explanation: If the goal is to provide an explanation of Odile's response with regard to an occurrence, then the explanation is best couched in terms of Odile's belief about that occurrence. Note that, intuitively, Odile's believing that (O1) express a truth is both necessary and sufficient for her being disposed to sincerely assent to (O1), while Odile's holding the attitude of belief toward the content expressed by (O1) is only necessary for her being so disposed: she may hold the attitude of belief toward the content expressed by (O1), and yet not know what content is expressed by (O1).)

Why might it seem that intentional Explanation B is false under a Russellian interpretation? Part 1 is unproblematic for the Russellian: if lines (i) and (ii) are true, then, *ceteris paribus,* line (iii) must also be true. Part 2, however, seems problematic for the Russellian interpretation. The seeming difficulty arises because, under the Russellian interpretation, it is not true that Odile does not BELIEVE that Clemens is a great author. Notice, however, that the truth of lines (i) and (ii) of part 2 do not entail that Odile does

239

not BELIEVE that Clemens is a great author. Line (i) of part 2 is true under the Russellian interpretation just in case Odile DESIRED that she indicate that she did not BELIEVE that Clemens is a great author. But Odile can DESIRE that she indicate that she did not BELIEVE that Clemens is a great author, even if she did in fact BELIEVE the content that Clemens is a great author. (This DESIRE of Odile's is similar to my DESIRE that I be given a prize for winning the Boston Marathon--the fact that I came in more than two hours behind the winner does not imply that I have no such DESIRE.) It is perfectly compatible with the Russellian view that Odile both hold the attitude of belief toward the Russellian proposition that Clemens is a great author, and also hold the attitude of belief toward the Russellian proposition that she does not believe the content that Clemens is a great author. Consequently, line (i) of part 2 is true under the Russellian interpretation. But if the Russellian theorist can endorse line (i), then lines (ii) and (iii) are unproblematic.

I conclude that Russellian approach is just as amenable to a comprehensive intentional psychological theory as the Fregean approach, and therefore that the allegation that Fregean theories are better suited to account for the cognitive significance of occurrences is unfounded.

What, then, accounts for the common allegation that Fregean theories are more psychologically plausible than Russellian theories? I suggest this allegation arises out of a conflation of semantic and psychological, or epistemological, issues. It is certainly correct that our untutored intuitions dictate that appropriate occurrences of sentences such as

(38) Odile believes that Twain is a great author, but not that Clemens is a great author, and that's why she assented to (O1), yet dissented from (O2).

are true. According to the Naive Russellian Theory, however, such occurrences are false, and it is a short, though unwarranted, step from this *semantic* fact to the epistemological\psychological conclusion that the Russellian view cannot account for the cognitive significance of occurrences. I agree that the semantic component of an acceptable Russellian theory must have it that many occurrences of sentences such as (38) are true. But this semantic task should be distinguished from the epistemological and psychological task of accounting for the cognitive significance of occurrences. (The semantic task is taken up in section 4.4.)

4.3 The Appropriate Russellian Response to the Semantic Arguments From Opacity.

It is agreed on all sides that an occurrence of an attitude ascription may convey more information than is encapsulated in the Russellian proposition presented by the occurrence. This much is implied by the fact that we may intuit that an occurrence of

(3) Odile believes that Twain is a great author.

is true, while an occurrence of

(4) Odile believes that Clemens is a great author.

is false. Beyond this point, however, there is little agreement. For those who endorse the Principle of Propositional Truth, there are two options:

> Option (i): Accept as veridical our untutored intuitions concerning the truth values of the occurrences, and deny that the occurrences present the same proposition. (This is the option taken by Frege, Richard, and Crimmins and Perry.)

Option (ii): Deny the veracity of our untutored intuitions, and maintain that the occurrences present the same proposition. (This is the option taken by Salmon.)

Option (i) is untenable. Any plausible way of pursuing option (i) will involve an appeal to some kind of mediator--senses, sentences of mentalese, etc.,--and I have shown that such entities cannot be responsible for ordinary speaker's intuitions concerning the truth conditions of occurrences of ascriptions such as (3) and (4). Option (ii) is superior to option (i), but because it assumes the Principle of Propositional Truth, option (ii) denies the veracity of our untutored intuitions concerning the truth conditions of occurrences, and this is unacceptable. In this section I will argue in support of a third option which preserves the veracity of our untutored intuitions, yet denies the Principle of Propositional Truth. That is, in this section I will argue in support of the following option:

Option (iii): Maintain that occurrences of ascriptions such as (3) and (4) present the same proposition, and also preserve the veracity of our untutored intuitions, but reject the Principle of Propositional Truth.

Note that if this third option is accepted, the semantic arguments from opacity against Russellian theories are undermined. The third option denies the Principle of Propositional truth, and if this principle is denied, then it does not follow from the fact that occurrences of

(3) Odile believes that Twain is a great author.

and

(4) Odile believes that Clemens is a great author.

present the same Russellian proposition that these occurrences must have the same truth conditions. I am therefore suggesting that the appropriate response to the semantic arguments from opacity is to deny the soundness of these arguments by denying the Principle of Propositional Truth. I concede, however, that in denying the Principle of Propositional Truth, I take upon myself the burden of explaining what, if not simply the proposition presented, does determine the truth conditions of an occurrence. Thus in the remainder of this section I will sketch an account the truth conditions of occurrences which is in accordance with a Russellian view of propositions. (My views on these matters have been heavily influenced by work by Stalnaker, Kamp, and Heim.[20])

I have been assuming throughout that *occurrences* of sentences, as opposed to sentences, present propositions, where occurrences are particular utterances or instances of inscriptions in particular *contexts*. Up until this point, however, I have said next to nothing concerning what *contexts* are and how they contribute toward what is presented, and otherwise conveyed, by token sentences. The time has come to say something about contexts, and the role they play in determining the truth conditions of occurrences.

Let us define a *belief set* for a subject x at a time t as the set of Russellian propositions believed by x at t. (This set may, and probably does, contain incompatible propositions.) I assume that the *belief set* for a subject x at a time t is determined by the belief state of x at t, where a *belief state* is a "wide" state; i.e. full specification of a subject's *belief state* would require a detailed description of the subject specifying both internal properties, and certain relational properties which hold of the subject in virtue of causal relations between the subject and his environment. Let us now define the *context* for x and y at t as the set of Russellian propositions which are *mutually believed* by x and y at t. (Two subjects x and y *mutually believe* p just in case at t x and y believe p, and at t x and y believe that at t x and y believe p, and at t x and y believe that at t x and y believe

that at t x and y believe p, etc.) Thus the *context* for x and y at t is a subset of the intersection of the *belief sets* of x and y at t.

The central assumption of this model of communication is that the purpose of making an assertion is to alter the context relative to which the assertion is made, and thereby alter a listener's belief set. What is *communicated* by an occurrence is what, on the basis of that occurrence, the speaker adds to the context. (For the sake of simplicity, I assume throughout that all occurrences are *assertions*. This is a gross simplification, but a discussion of *force*, and other kinds of speech acts would, at this point, only complicate matters.) Upon correctly interpreting a speaker's utterance, a listener at least temporarily adds what is communicated to his belief-set, and derives various propositions as consequences. The derived propositions may or may not be acceptable, and on the basis of the acceptability of the derived propositions, the listener may or may not accept all, or part, of what is communicated into the context. (If the listener does not accept all or part of what is communicated, then, in a paradigmatic discourse, the listener must inform the speaker of what parts of what is communicated he does not accept. That is, there is a *presumption of acceptance*: if speaker x communicates p, and listener y does not protest, then x believes that y believes p, and y believes that x believes that y believes that p, etc..) Note that in asserting Russellian proposition p, x *communicates* to y much more than the *presented* Russellian proposition p; in asserting p to y, x communicates to y a good deal of "pragmatically imparted" information. For example, one crucially important piece of information that is communicated by x's assertion of p is the proposition that x holds the attitude of belief toward p; let us call this proposition which is communicated by x's assertion of p the "Moore proposition for x's assertion of p." (The fact that an assertion of p communicates a Moore proposition accounts for the contradictory feel of occurrences of "Moore's Paradox" sentences; e.g. "Twain is a great author, but I don't believe it." An occurrence of the first conjunct *communicates* that Twain is a great author and that the speaker believes that Twain is a great author, but this

latter proposition is contradicted by the proposition *presented by* the second conjunct. Hence an occurrence of the sentence communicates incompatible propositions.) It follows from the above claims (coupled with a few other assumptions) that, in paradigmatic discourse, if x asserts p to y and y does not protest, then p is mutually believed, and thus is added to the context of the discourse. (Roughly, because x's assertion of p communicates both p and the Moore proposition that x believes that p, a listener y who accepts x's assertion comes to believe both p and that x believes p. And because listener y does not protest x's assertion of p, x comes to believe that y also believes p, and that x believes p. Finally, because x and y are competent speakers and believe of one another that they are competent speakers, they come to have mutual belief in p; when listener y accepts p and that x believes p, both x and y come to believe not only p and that x believes p, but also that they share these beliefs, and that they share the belief that they share these beliefs, etc.)

According to the model of communication I am dogmatically sketching here, ordinary discourse is a process whereby two or more subjects cooperatively amend a context. In most cases this process is goal directed; in most cases, the participants of a conversation are cooperatively attempting to perform some task, answer some question, or increase their knowledge concerning some topic, and that they are attempting to do this is also mutually believed. In other words, in paradigmatic discourse the participants are cooperatively attempting to construct a context which meets some general description. This fact is often exploited in the process of communication; the fact that the participants share a desire to construct a context of a certain kind can be exploited to resolve ambiguities. For example, if our goal in a discourse is to decide whether or not a person should be allowed to fly a plane, then it is clear how my utterance of 'Flying planes can be dangerous' is to be interpreted. And, more to the point, it is the mutually understood goals of a discourse which determine whether or not a context is *sensitive* or *insensitive*; i.e. it is the mutually understood goals of the discourse which determine

whether or not an ascription is to be interpreted *de re* or *de dicto*. (This point will be clarified, to some extent, below.)

Contexts are what are constructed through processes of ordinary discourse. Contexts also serve, however, an essential function within the processes of their own construction; contexts also serve as what Kaplan called "contexts of utterance." In uttering Σ in order to assert p to y, x presupposes that y has the requisite knowledge, or beliefs, for correctly interpreting the utterance of Σ; e.g. if x uses the name 'Twain', x presupposes that y is familiar with this name, and if x uses 'he', then x presupposes that y knows, or can determine, who the intended referent is. This implies that speakers must-- if they want to be understood--limit their utterances to utterances of sentences which require for correct interpretation only information which is already in the context. (In special cases of *accommodation* an utterance U may itself introduce information into the context which is necessary for the interpretation of U, but I will ignore these special cases here.[21]) Note, however, that as a conversation progresses and the context is amended, the abilities of the participants to interpret utterances correctly will change, and as a result the limitations on what the participants are licensed to utter also changes. Whether or not a participant is licensed to utter a sentence Σ--whether or not he can expect his utterance to be interpreted correctly--depends in part upon previous utterances in the discourse. (E.g. one might not be able to begin a sentence with 'He' unless someone has already attempted to refer to the intended referent of 'He'.) The context is thus not only what the participants in a discourse construct and amend, it also places constraints upon the sentences that can be uttered in order to amend it.

Given this model of communication, what determines the truth conditions of *an occurrence*? A Russellian proposition, either presented or otherwise conveyed, is true just in case it corresponds with the actual world, just in case it *obtains*. But what is it for *an occurrence*--which typically *communicates* many propositions--to be true? If the truth conditions of *the proposition presented* by an occurrence of an attitude ascription do not

determine the truth conditions of *the occurrence*, then what does determine the truth conditions of an occurrence? In what follows I first give a very rough answer to this question, and then give a slightly more precise answer in terms of the "Discourse Representation Theory" proffered by Kamp.[22]

The very rough answer is that Salmon is more-or-less correct with regard to what is responsible for our intuitions concerning the truth conditions of occurrences: more information is conveyed by an occurrence than the proposition presented by the occurrence, and our intuitions concerning the truth conditions of occurrences are sometimes sensitive to this additional information. Unlike Salmon, however, I do not think that ordinary speakers are systematically confused concerning the truth conditions of occurrences in their own language; ordinary speakers do not judge incorrectly when they judge that an occurrence of

(4) Odile believes that Clemens is a great author.

uttered in a sensitive context, is false. But what is the relevant "extra" information which is conveyed by an occurrence of an attitude ascription in a sensitive context which accounts for our judgments of truth conditions? Consider the following situation: You are in a conversation with Odile, and she asserts

(1) Twain is a great author.

Later, you are in a conversation with Oscar. Oscar asks you, "Does Odile believe that *Clemens* is a great author?" Even if you have good reason to suppose that Odile has not "changed her mind," you are not warranted to reply affirmatively. You would, however, be warranted to respond affirmatively if Oscar had instead asked you, "Does Odile

believe that *Twain* is a great author?" I suggest that this implies that the information

responsible for our differing intuitions with regard to

(3) Odile believes that Twain is a great author.

and

(4) Odile believes that Clemens is a great author.

is conveyed by Odile's utterance of (1). That is, Odile's utterance of (1) pragmatically

conveys information whose truth suffices for the truth of occurrences of (3), but not

occurrences of (4).

So what is the relevant information that Odile's utterance of (1) pragmatically

conveys? What proposition conveyed by Odile's utterance of (1) warrants, in a sensitive

context, a judgment that an occurrence of (3) is true, yet does not warrant a judgment that

an occurrence of (4) is true? Odile's utterance of (1) pragmatically conveys much

information. One of the propositions it conveys is the proposition the occurrence

presents, viz. that Twain (the man) is a great author. But Odile's utterance also conveys

the Moore proposition for her utterance, and other propositions which are similar to this

Moore proposition. For example, Odile's utterance of (1) conveys that Odile believes

that 'Twain is a great author', relative to the context in which it is uttered, presents a true

proposition. Thus, Odile's utterance conveys, among other things, that she thinks 'Twain'

relative to c, has a referent, and that this referent is a great author. (Note the paradoxical

quality of utterances of sentences such as, "Twain is a great author, but I don't think

'Twain' refers to a great author.") I suggest that it is this information concerning Odile's

beliefs concerning 'Twain' and what it refers to which is responsible for the differing

truth conditions of occurrences of (3) and (4): An occurrence of (3), in a sensitive

context, *communicates*--adds to the context in which it is uttered--the proposition that Odile believes that 'Twain' has a referent who is a great author, and an occurrence of (4), in a sensitive context, communicates the proposition that Odile believes that 'Clemens' has a referent who is a great author. Thus, assuming that Odile does not believe that 'Clemens' refers to a great author, to utter (4) in a sensitive context would be to add false information to the context.[23]

What about truth conditions? The proposition *presented by* occurrences of (3) and (4) is true just in case it is actualized; i.e. just in case Odile holds the attitude of belief toward the Russellian proposition that Twain is a great author. *Occurrences* of (3) and (4), on the other hand, have, when uttered in sensitive contexts, distinct truth conditions: the truth of an occurrence of (3) may require that Odile believe that 'Twain' has as its referent a person named 'Twain', while an occurrence of (4) instead requires that Odile believe that 'Clemens' has as referent a person named 'Clemens'. But how is this proposal to be accommodated under the model of communication adumbrated above? The idea is roughly this: the truth value of an occurrence is determined by comparing the propositions the occurrence adds to context c with the actual environment: an occurrence in a context c which brings about context c' is true just in case the propositions which the occurrence adds to c correspond with the actual world. Under this proposal the truth value of an occurrence is not merely a matter of what the occurrence *presents*, but rather concerns what effect the occurrence has upon the context in which it occurs. Consequently, an occurrence of (4) may be false in a context c because relative to c an utterance of (4) would communicate, among other things, the proposition that Odile believes that 'Clemens' refers to a great author. This proposition, however, does not correspond with the actual world, and thus the occurrence is false. (Thus a *sensitive context* for (4) is a context such that the proposition that Odile believes that 'Clemens' refers to a great author is communicated by an occurrence of (4) relative to that context.[24] But what determines whether or not an occurrence of (4) communicates

this proposition? I cannot give an adequate answer to this question, but for present purposes it will suffice to say that whether or not this proposition is communicated is determined by the shared goals of the participants of the discourse.)

This sketch of what it is for an occurrence to be true is, I think, roughly correct, but it does not suit my purpose of motivating option (iii), described above. According to option (iii), the judgments ordinary speakers make concerning the truth conditions of occurrences of attitude ascriptions are *correct*; e.g. ordinary speakers are correct in judging that an occurrence of (4), in a sensitive context, is false. Hence, in order to motivate option (iii), I require an account of what it is for an occurrence to be true which not only *accords with* our untutored intuitions concerning the truth conditions of occurrences, but also *explains* how and why ordinary speakers have the intuitions and make the judgments they do; in order to motivate option (iii) properly, I require an account of truth for occurrences which at least helps to explain how it is that speakers are *able to know* the truth conditions of occurrences. The account of truth conditions of occurrences sketched above *accords with* our intuitions concerning truth conditions, but, as it is stated in terms of a correspondence between contexts--sets of Russellian propositions--and the actual world, it tells us nothing concerning how ordinary subject's are able to correctly judge of an occurrence that it is true (or false.) The account sketched above does nothing to explain the correspondence between the intuitive judgments of ordinary speakers and the truth conditions of occurrences. If I am to motivate option (iii), I must give an account of the truth conditions of occurrences which is rooted in psychological reality; I must give an account which shows and explains why an omniscient subject--a subject whose belief set contained all and only the Russellian propositions actualized in the world--would make judgments in accordance with our untutored intuitions. In the remainder of this section I will illustrate how an account of truth for an occurrence which at least attempts to satisfy this psychological requirement can be given by *Discourse Representation Theory*, or DRT.

A context for x and y at t is a set of Russellian propositions. This set, however, is determined by the *belief states* of x and y at t (where belief states are "wide" states). Hence a change in the context of x and y must be accompanied by, and due to, a change in the belief states of x and y; as the context of a paradigmatic discourse is amended, there will be corresponding alterations in the belief states of the participants. Thus, in a paradigmatic discourse with participants x and y, there will be a kind of correspondence between parts, or features, of the belief states of x and y. Discourse Representation Theory is, I propose, best thought of as a formal means of representing this dynamic correspondence between the belief states of the participants of a discourse. A Discourse Representation Structure (or DRS) is a model of the shared features of the belief states of the participants in a successful discourse. (A *discourse* is chronologically well-ordered finite sequence of well formed utterances. A *paradigmatic discourse* is a discourse which proceeds via certain rules which facilitate mutual understanding among the participants. One of the goals of DRT is to discover, and formulate, the rules governing paradigmatic discourse.) Because DRT is concerned first and foremost with *belief states* rather than the propositions those states determine, DRT is well suited to satisfy the psychological requirement discussed above. That is, since DRT is a formal means of representing the correspondence between the *belief states* of two or more participants of a discourse, an account of the truth conditions of occurrences in terms of discourse representations will be well suited to explain the correspondence between the judgments of ordinary speakers and the truth conditions of occurrences of attitude ascriptions.

The best way to introduce DRT is to illustrate how it applies to a particular example. Consider a discourse which consists of utterances of the following sentences:

Hob: The witch poisoned Spot.
Nob: She poisoned Rover too.[25]

Hob's utterance occurs within a certain context. That is, Hob and Nob mutually believe a set of Russellian propositions at the outset of the discourse. The belief states of Hob and Nob, which both determine this context, will therefore determine an initial discourse representation structure; this discourse representation structure *characterizes* the belief states of Hob and Nob. (What it is for a DRS to *characterize* a mental state will be defined below.) A plausible initial DRS for this discourse is depicted by the following diagram:

DRS-1a

Universe: *a, b, c,*

Conditions: 1. *a* is a witch
2. *b* is named 'Spot'
3. *b* is a dog
4. *c* is named 'Rover'
5. *c* is a dog

A few points of clarification concerning DRS-1a are in order. First, the symbols '*a*', '*b*' etc. do not refer to ordinary entities "in the world." Rather they refer to *reference markers*, where a reference marker is a mental representation type. E.g. '*a*' does not, of course, refer to a certain witch. Rather '*a*' refers to a type, tokens of which are the mental representation tokens which allow Hob and Nob to seemingly speak and think about the *same* witch. (Thus a reference marker may fail to represent anything.) Nor, strictly speaking, do the predicates appearing in the conditions designate properties. Rather strictly speaking the predicates also designate mental representation types. E.g. the predicate 'is a dog' appearing in condition 3 does not designate the property *being a dog*, but instead designates a type, tokens of which are the mental representation tokens which allow Hob and Nob to seemingly speak and think about the same property, the property *being a dog*. For the sake of simplicity, however, in what follows the predicates

appearing in conditions can be interpreted as serving double duty and designating both properties and the relevant mental representation types. (Since properties are easily purchased, predicates are almost guaranteed to have designations. Thus each of the predicates appearing in the conditions designates a mental representation type which, in turn, is guaranteed to represent a property.) The symbols appearing in the above depiction of DRS-1a represent correspondences between the belief states of Hob and Nob, and nothing more. Thus DRS-1a can be interpreted as representing the fact that Hob and Nob are in belief states such that *if* there were real entities corresponding to '*a*', '*b*' and '*c*', then conditions 1-5 would represent Russellian propositions in the context of Hob and Nob's discourse.

Second, DRS-1a is highly impoverished. A more plausible initial discourse representation structure for Hob and Nob's discourse would involve a vastly larger number of elements in the universe of the DRS, and many more conditions. In what follows, I will attempt to keep the DRSs as simple as possible, but it is to be understood that the DRSs posited by a complete theory of discourse representation would be much more complex.

Third, note that there is an entity in the universe of DRS-1a for each definite NP appearing in Hob's utterance, the first occurrence of the discourse. This is as required by the following general rule for DRS construction:

> The Familiarity Constraint: Every definite NP appearing in an occurrence *corresponds with* a reference marker which is *linked to* a reference marker already present in the DR.

(What it is for an occurrence of an NP to *correspond with* an already present reference marker will be made clear below. For now, let it be understood that every occurrence of an NP, whether definite or indefinite, *corresponds with* a reference marker. What is it for reference markers m and m' to be *linked*? Roughly, the linked relation is an

equivalence relation, and reference markers *m* and *m'* are linked just in case *m* and *m'* *correspond with* occurrences NP1 and NP2 respectively, and NP1 and NP2 are assumed by the participants of a discourse to corefer or codenote. Thus the *corresponds with* relation holds between occurrences of NPs and reference markers, while the *linked to* relation holds between reference markers. Of course, as the *linked to* relation is an equivalence relation, every reference marker is linked to itself. These terms and conditions will be clarified as we proceed.) The intuitive idea behind The Familiarity Constraint is that occurrences of definite NPs *presuppose* that all participants of a discourse understand these NPs and believe that they have referents. (Cases of accommodation are exceptions to this general constraint.) Hob's utterance, for example, presupposes that Hob and Nob mutually believe that there is some unique salient witch, and that there is some unique salient dog named 'Spot'. Furthermore, Hob assumes that Nob will be able to "connect" these definite NPs with these existential beliefs; Hob assumes that Nob will be able to determine that Hob's utterance of 'Spot' refers to Spot, the dog they both know and love. There is, of course, an important difference between the two NPs; viz. the NP 'the witch' denotes nothing, and thus the reference marker corresponding to this NP, reference marker *a*, does not represent an entity "in the world." The NP 'Spot', we may suppose, has a referent and thus the reference marker corresponding to this NP, reference marker *b*, does represent an entity "in the world." This semantic difference between the two NPs, however, does not arise at the level of discourse representation structure.

Associated to the Familiarity Constraint for definite NPs there is another constraint which applies to indefinite NPs:

> The Novelty Constraint: Every indefinite NP appearing in an occurrence *corresponds with* a reference marker which is not *linked to* a reference marker already present in the DR.[26]

When a speaker uses an indefinite NP, he typically assumes that his listener is not familiar the individual denoted by the NP. For instance, if I report to you, "I met *a woman*," unless I am being coy, I am assuming that you are not familiar with the woman I met. If I thought you were familiar with her, and that you knew that her name was 'Mary', I would have instead reported to you, "I met Mary." (If there were no such name, I would probably attempt denote her via a definite description, e.g. "the woman you spoke to when . . .")

Discourse is a dynamic process whereby a context is amended. Since a change in context can only occur if there are corresponding changes in the belief states of the participants, discourse is also a dynamic process whereby a DRS is amended. That is, Hob's utterance of 'The witch poisoned Spot' induces a change in DRS-1a. Allowing for gross simplification, it may be supposed that Hob's utterance relative to DRS-1a induces the following DRS:

DRS-1b

    Universe: *a, b, c,*

    Conditions: 1. *a* is a witch
                    2. *b* is named 'Spot'
                    3. *b* is a dog
                    4. *c* is named 'Rover'
                    5. *c* is a dog
                    6. *a* poisoned *b*

Hob's utterance does not introduce any new elements into the universe of DRS-1a. This is because Hob's utterance contains definite NPs only, and thus the reference markers *corresponding with* the definite NPs 'the witch', and 'Spot'--*a* and *b* respectively--must be *linked to* reference markers already present in the universe of DRS-1a. How does DRS-1b indicate that reference marker *a* corresponds with the occurrence of 'the witch', and

that reference marker *b* corresponds with the occurrence of 'Spot'? The only difference between the sentence uttered by Hob and condition 6--the one condition added to the conditions of DRS-1a by Hob's utterance--is that condition 6 has the reference marker symbols '*a*' and '*b*' instead of the NPs 'the witch' and 'Spot', respectively. This indicates that the definite NP 'the witch' *corresponds with* reference marker *a*, and the NP 'Spot' *corresponds with* reference marker *b*. Consequently, since all reference markers are linked to themselves, and both reference marker *a* and reference marker *b* are in the universe of DRS-1a, it follows that the reference markers corresponding to the occurrences of 'the witch' and 'Spot' are *linked to* reference markers already in the universe of the DR. Thus the Familiarity Constraint for definite NPs is satisfied. (Note that if Hob had instead uttered, '*A witch* poisoned Spot', then either reference marker *a* does not allegedly represent the witch Hob is seemingly talking about, or DRS-1a is not an appropriate initial DRS for the discourse.)

But why is that the occurrences of 'the witch' and 'Spot' are made to correspond with reference markers *a* and *b*, respectively? Why not, for instance, make the occurrence of 'the witch' correspond with reference marker *b*, and the occurrence of 'Spot' correspond with reference marker *c*? A full fledged theory of discourse representation would state precisely how such correspondences are determined, but here I will simply assume that the most intuitively plausible correspondences and linkages are correct. Thus, since condition 1 indicates that reference marker *a* allegedly represents an individual who is a witch, and conditions 2 and 3 indicate that reference marker *b* allegedly represents a dog named 'Spot', the most plausible correspondence matches the NP 'the witch' with reference marker *a*, and the NP 'Spot' with reference marker *b*.

Finally, let us consider how Nob's response to Hob's utterance affects DRS-1b. Nob's utterance relative to DRS-1B induces DRS-1c:

DRS-1c

Universe: *a, b, c,*

Conditions: 1. *a* is a witch
2. *b* is named 'Spot'
3. *b* is a dog
4. *c* is named 'Rover'
5. *c* is a dog
6. *a* poisoned *b*
7. *a* poisoned *c*

Nob's response adds condition 7 to DRS-1b. Because 'she' is a definite NP, it cannot introduce a reference marker into the universe of DRS-1b, but rather must correspond with a reference marker which is linked to a reference maker already present in DRS-1b. What determines which one of the already present elements this occurrence of 'she' corresponds with and is linked to? In particular, why does the occurrence of 'she' correspond with reference marker *a*, and not to reference marker *b*? This is a complicated issue, and I will not discuss it any detail here. Suffice it to say that the adverb 'too' appearing in Nob's utterance determines that 'She' can correspond with, and be linked to, reference marker *a* only. (Note that if Nob had instead uttered, 'She poisoned Rover', the occurrence of 'she' could have corresponded with, and been linked to, either *a* or *b*.)

What about truth conditions? What is it for Nob's response--*the occurrence*--to be true? Very roughly, Nob's response is true just in case there is some way of assigning actual entities, entities "in the world," to the reference markers of the universe of DRS-1c which satisfies all the conditions of DRS-1c. That is, Nob's utterance is true just in case there is a (partial) *assignment function f*( ) from discourse referents to actual entities such that, 1. $f(a)$ is a witch; 2. $f(b)$ is named 'Spot'; 3. $f(b)$ is a dog; 4. $f(c)$ is named 'Rover'; 5. $f(c)$ is a dog; and 6. $f(a)$ poisoned $f(b)$. (I will say that Nob's utterance is true relative to DRS-1b just in case there is an *appropriate assignment function f*( ) such

that conditions 1-7 are *satisfied under assignment function f( )*. An assignment function *f( )* is *appropriate for* a DRS D just in case *f( )* is defined for all and only the elements of the universe of D.) Note that for an occurrence to be *false* is not simply for there to be no appropriate assignment function. If there is no appropriate assignment function, then the occurrence is *not true*; i.e. it is either false, or infelicitous. (This issue will be discussed in more detail below.)

This definition of truth for occurrences, however, does not accord with our untutored intuitions concerning the truth condition of occurrences. The problem is that the above definition does not seem to capture our intuitions concerning the truth conditions of occurrences containing definite descriptions, nor occurrences containing directly referential terms. Our untutored intuitions dictate that Nob's utterance is true only if there is some *particular* witch, the witch whom Hob "has in mind," and this particular witch poisoned *the very dog Spot*. According to the definition of truth for an occurrence stated above, however, it is sufficient for the truth of Nob's utterance relative to DRS-1b that some witch or other poisoned some dog or other named 'Spot'. Thus the above account of truth must be amended.

On reflection, it is not at all surprising that the above description of DRSs coupled with the above definition of truth for an occurrence fail to capture our intuitions concerning the truth of occurrences. Our intuitions concerning the truth conditions of an occurrence are sensitive to the referents of the terms appearing in the occurrence and, as Putnam and Burge have taught us, the referent of an occurrence cannot be determined by the "narrow" mental state of the speaker alone. As thus far described however, DRSs are models of narrow mental states only, and thus it is not surprising that the above definition of truth for an occurrence does not capture our intuitions concerning reference. If DRSs and something like the above definition of truth for an occurrence are to capture our untutored intuitions concerning the truth conditions of occurrences, DRSs must be augmented so that they somehow specify appropriate world-mind relations. That is,

something must be added to DRSs to account for that feature of reference determination which occurs "outside of the head," and the above definition of truth must be amended accordingly. (If predicates are "directly designating" terms, a similar difficulty applies to predicates. Again, for the sake of simplicity, this issue will be overlooked here.)

Consider the occurrence of 'Spot' which appears in Hob's utterance. What is needed is some way of ensuring that the reference marker corresponding with this occurrence, viz. reference marker b, is mapped to the very dog Spot. More specifically, DRS-1b and the definition of truth for an occurrence must be amended so that Hob's utterance is true only if there is some assignment function $f(\ )$ such that $f(b) = $ Spot. This can be accomplished by adding a set (perhaps the empty set) of *external restrictions* to DRSs. One *external restriction* is added to a DRS for every successfully referring directly referential NP of a discourse. (I will say that a restriction *governs* the reference markers occurring in it. Later *internal restrictions*--which determine what reference markers are linked--will also be added to DRSs, and these internal restrictions will also be said to *govern* the reference markers occurring in them.) For example, DRS-1b, the DRS which is induced by Hob's utterance relative to DRS-1a, is amended as follows:

DRS-1b'

Universe: *a, b, c,*

Conditions: 1. *a* is a witch
2. *b* is named 'Spot'
3. *b* is a dog
4. *c* is named 'Rover'
5. *c* is a dog
6. *a* poisoned *b*

External Restrictions: 1. R(*b*) = Spot

External restriction 1 governs reference marker b; it states that the reference marker b represents Spot, the very dog referred to by Hob's utterance. There is no external restriction for reference marker a because, first, the occurrence of 'the witch' of course lacks a denotation, but more importantly the NP 'the witch' is not a directly referential NP; hence *even if* the occurrence of 'the witch' had a denotation, no external restriction would be added to DRS-1b'. ('R( )' designates a partial function from reference markers to the actual entities they represent, if they represent an actual entity. I will say that an assignment function $f( )$ *obeys* an external restriction governing a reference marker x just in case $f(x) = R(x)$.)

What about the (empty) definite description 'the witch' that appears in Hob's utterance? Here matters are more complicated. The difficulty concerns the *presupposition of uniqueness* for definite descriptions. Our intuitions dictate that Hob's utterance is true just in case there is some unique witch who poisoned Spot, but under the definition of truth given above, Hob's utterance is true just in case some witch or other poisoned Spot. How are we to solve this problem? One suggestion would be to add external restrictions governing the reference markers corresponding to definite descriptions. Assume for a moment that there are witches, and that one of them, say Hilda, poisoned Spot. Under this assumption, our intuitions seem to indicate that Hob's utterance is true just in case Hilda poisoned Spot, and we could ensure that our account of truth for an occurrence captured this intuition by adding the external restriction, $R(a) =$ Hilda, to DRS-1b. While this proposal would solve the problem, I do not think it is in the spirit of a Russellian view: A cornerstone of the Russellian point of view is the idea that definite descriptions and directly referential terms function in different ways, and the above proposal suggests that they be treated in more or less the same way. Hence I will reject the suggested proposal in favor of what I think is a more illuminating solution to the problem.

As a way of motivating the solution I will offer, I want to highlight some often overlooked similarities between definite and indefinite descriptions. Consider an occurrence of 'The fattest man weighs over 400 lbs.' In a typical utterance of this sentence, the speaker will not have a particular man in mind, and the speaker will not be attempting to communicate something about a particular man. Rather a typical occurrence of this sentence will be interpreted as communicating something like "the fattest man, *whoever he is*, weighs over 400 lbs." But now consider a typical utterance of 'The fat man hit me.' Here, it seems, the speaker has a particular person in mind, and his utterance is true just in case that particular person is a fat man who hit the speaker. Thus sometimes the utterer of a definite description "has someone in mind," and sometimes the speaker does not "have someone in mind." The same distinction applies to occurrences of sentences containing indefinite descriptions. Consider an occurrence of 'A fat man must have sat here.' Here again in a typical utterance of this sentence, the speaker would have no particular man in mind; the utterance would be true just in case at least one fat man sat in the place indicated. Compare the former with a typical utterance of 'A fat man hit me.' Here again, the speaker would typically have someone in particular in mind. Thus the difference between definite and indefinite descriptions is not that definite descriptions are used to "single out" a particular entity, while indefinite descriptions are merely used to quantify over and describe entities; both definite and indefinite descriptions are used to "single out" entities, and to quantify over and describe entities.

What then is the difference between definite and indefinite descriptions? The difference concerns truth conditions. It is not the case that an occurrence of 'The fat man sat here' is true just in case at least one fat man sat in the place indicated. But it is the case that an occurrence of 'A fat man sat here' is true just in case at least one fat man sat in the place indicated. Definite descriptions, but not indefinite descriptions, *presuppose uniqueness*. But what does this mean? It is traditionally taken to mean that an occurrence of a definite description carries a presupposition to the effect that the

261

occurrence must somehow manage to "single out" one entity from the domain of all the entities there are. (Never mind how an occurrence might do this, and whether or not this idea is coherent.) I suggest, however, a much different way of understanding the presupposition of uniqueness carried by occurrences of definite descriptions. I suggest that a felicitous occurrence of a definite description need only "single out" one of the reference markers in the DRS relative to which the definite description was uttered. Recall that in order to satisfy the Familiarity Constraint, the reference marker corresponding with an occurrence of a definite description must be linked to an already present reference marker. What I am suggesting is that the presupposition of uniqueness carried by occurrences of definite descriptions concerns the *linked to* relation, and not the *denotes* relation. For example, I maintain that an utterance of 'the fat man sat here' is felicitous--satisfies the presupposition of uniqueness--relative to a DRS just in case there is only one reference marker *m* to which the reference marker corresponding with the definite description may be appropriately linked. The fact that there are many fat men in the world is irrelevant to the felicity of the occurrence. If this way of understanding the presupposition of uniqueness carried by definite descriptions is correct, then nothing must be added to DRSs in order to account for our intuitions concerning the truth conditions of occurrences of sentences containing definite descriptions. Under my proposal the presupposition of uniqueness is a consequence of the definiteness of definite descriptions, and is therefore (already) enforced by the Familiarity Constraint.

Perhaps an example can help to clarify this point. Suppose Oscar and Otho are looking at a broken chair, and engage in a discourse. Suppose that the initial DRS for their discourse is as follows:

<u>DRS-2a</u>

Universe: *a*

Conditions: 1. *a* is a man
2. *a* is fat

Oscar now utters, pointing to the broken chair, an occurrence of

(39) The fat man sat here.

This utterance induces the following DRS:

<u>DRS-2b</u>

Universe: *a*

Conditions: 1. *a* is a man
2. *a* is fat
3. *a* sat here (the place indicated).

Thus the occurrence of the definite description 'The fat man' corresponds with reference marker *a*, and, since reference marker *a* is already present in DRS-2a, Otho's utterance is in accordance with the Familiarity Constraint. What about the presupposition of uniqueness? I maintain that the presupposition of uniqueness is satisfied because there is only one reference marker in DRS-2a with which the occurrence of 'the fat man' can appropriately correspond. (Again, a complete theory of discourse representation would not depend upon intuitions at this crucial juncture, but would instead offer a theory explaining why the occurrence of 'the fat man' must correspond with reference marker *a*.) Since there is no ambiguity concerning which already existing reference marker 'the fat man' is to correspond with and be linked to, Oscar's utterance is felicitous. In contrast,

suppose that the initial DRS for Oscar and Otho's discourse was not DRS-2a, but rather as follows:

DRS-2a'

Universe: *a, b*

Conditions: 1. *a* is a man
2. *a* is fat
3. *b* is a man
4. *b* is fat

The presupposition of uniqueness for Oscar's utterance of (39) relative to DRS-2a' would not be satisfied, and consequently the occurrence would not be felicitous; i.e. Oscar's utterance of (39) does not induce a unique DRS, and thus does not have unique truth conditions. The problem is that it is indeterminate, or ambiguous, as to which already present reference marker the occurrence of 'the fat man' is to correspond with and be linked to: Oscar's utterance of (39) relative to DRS-2a' would be infelicitous because the occurrence of 'the fat man' appearing in Oscar's utterance could be appropriately taken as corresponding with either reference marker *a*, or reference marker *b*.

It might be objected that in interpreting the presupposition of uniqueness to concern the *corresponds with* relation, as opposed to the *denotes* relation, I have not succeeded in capturing our intuitions concerning the truth conditions of occurrences containing definite descriptions. An objector might argue as follows:

Suppose that DRS-2a is the initial DRS for Oscar and Otho's discourse, and that Oscar utters an occurrence of

(39) The fat man sat here.

and thereby induces DRS-2b. Further suppose, however, that unbeknownst to Oscar and Otho, there are two fat men, fm_1 and fm_2, in the vicinity, and that Oscar and Otho are only familiar with fm_1. Finally, suppose that fm_2, but not fm_1, actually sat in the chair and broke it. According to the proposed interpretation of the presupposition of uniqueness, Oscar's utterance is felicitous and true in case there is a assignment function $f(\ )$ such that $f(a)$ is a fat man who sat in the place indicated. Thus, since there is an assignment function $f^*(\ )$ such that $f^*(a) = fm_2$, the proposed interpretation of the presupposition of uniqueness predicts that Oscar's utterance is felicitous and true, but this prediction does not accord with our intuitions. Our intuitions dictate that Oscar's utterance is infelicitous because the presumption of uniqueness is not satisfied: the occurrence of 'the fat man' can be interpreted as denoting either fm_1 or fm_2, and this is why our intuitions dictate that Oscar's utterance is infelicitous.

The above objection assumes that in the described situation our intuitions dictate that Oscar's utterance relative to DRS-2a is infelicitous. My response to the objection is to deny this assumption: Our intuitions do not dictate that Oscar's utterance *relative to DRS-2a is infelicitous*, rather the above described situation elicits the intuition that if Oscar were to utter an occurrence of (39) relative to a DRS very much different from DRS-2a, then this *distinct* occurrence of (39) would be infelicitous. In DRT the truth conditions of an occurrence can be evaluated only relative to an initial DRS; an occurrence is a token of a declarative sentence uttered relative to an initial DRS. Relative to DRS-2a, Oscar's utterance of

(39) The fat man sat here.

is both felicitous and true. What the above objection implicitly asks the reader to do is to evaluate a distinct (imaginary) occurrence. More specifically, the above objection

implicitly asks the reader to evaluate an utterance of (39) not relative to DRS-2a, but instead relative to a DRS in which there are two reference markers, one corresponding with 'fm$_1$', and the other corresponding with 'fm$_2$'. In other words, in describing the situation in which there are two fat men, the objector makes the reader a participant in an imaginary discourse for which a DRS similar to DRS-2a' is the initial DRS. Relative to this newly introduced DRS, an utterance of (39) is infelicitous, but this is precisely what is predicted by my analysis of the presupposition of uniqueness. Relative to a DRS similar to DRS-2a' an utterance of (39) is infelicitous because the occurrence of the definite description can be appropriately taken to correspond with either of two already present reference markers. Thus the intuitions elicited by the above objection do not in any way conflict with the predictions offered by my account of truth conditions; they are just the intuitions that my account of truth conditions would predict the reader to have. Consequently the above objection does not pose a problem for my interpretation of the presupposition of uniqueness carried by occurrences of definite descriptions; indeed, the above objection offers evidence in support of my interpretation.

Truth for an occurrence can now be provisionally defined in accordance with our intuitions concerning directly referential terms and definite descriptions:

A token of sentence Σ uttered relative to DRS D which induces a DRS D' is true iff there is an appropriate assignment function $f(\ )$ such that (i) $f(\ )$ satisfies all the conditions of D', and (ii) $f(\ )$ obeys all internal and external restrictions which govern elements of the universe of D'.

(*Internal restrictions* are introduced into DRSs below.) The above definition is provisional because it is not intuitively adequate. The intuitive inadequacy of the above definition arises because the truth of an occurrence is defined in terms of DRSs, and DRSs are constructed out of finite sequences of occurrences. Consequently, according to the above definition, the truth of the nth occurrence of a discourse in part depends upon

how the 1st through the $(n-1)$th occurrences affected the DRS being constructed. For example, if the first occurrence in a discourse is an utterance of 'A witch poisoned Spot', then all subsequent occurrences in the discourse will *not be true*, as the condition added to the DRS by the first occurrence cannot be satisfied by any assignment function. This result is counterintuitive; the truth conditions of an occurrence should not *always* depend upon the truth conditions of all the occurrences preceding it. (Though *often* the truth conditions of an occurrence depend upon the truth conditions of the occurrences preceding it). I optimistically maintain that it is possible to refine DRT so that intuitively correct sufficient and necessary conditions for the truth of occurrences can be formulated, but I will not take on this difficult task here.[27] The provisional definition of truth for an occurrence stated above will suffice for our purposes.

Also note that it follows from the above definition that if there is no assignment function $f(\ )$ which satisfies conditions (i) and (ii) for an occurrence O which induces a DRS D, then O is *not true*; it does not follow that occurrence O is *false*. For example, consider Hob's utterance of 'The witch poisoned Spot' relative to

DRS-1a'

Universe: $a, b, c,$

Conditions: 1. a is a witch
2. b is named 'Spot'
3. b is a dog
4. c is named 'Rover'
5. c is a dog

External Restrictions: 1. $R(b) =$ Spot

(Note that DRS-1a' is DRS-1a with the addition of the appropriate *external restrictions*.) Hob's utterance adds condition 6 to DRS-1a', thereby inducing DRS-1b'

<u>DRS-1b'</u>

Universe: *a, b, c,*

Conditions: 1. *a* is a witch
2. *b* is named 'Spot'
3. *b* is a dog
4. *c* is named 'Rover'
5. *c* is a dog
6. *a* poisoned *b*

External Restrictions: 1. R(*b*) = Spot

Hence, according the above definition of truth for an occurrence, Hob's assertion is true if and only if there is an appropriate assignment function $f(\ )$ such that (i) $f(\ )$ satisfies conditions 1-6, and (ii) $f(\ )$ obeys external restriction 1. Putting all this together, Hob's utterance of 'The witch poisoned Spot' is true relative to DRS-1a' just in case (i) there is an assignment function $f(\ )$ such that 1. $f(a)$ is a witch; 2. $f(b)$ is named 'Spot'; 3. $f(b)$ is a dog; 4. $f(c)$ is named 'Rover'; 5. $f(c)$ is a dog; 6. $f(a)$ poisoned $f(b)$. And (ii), $f(b)$ is Spot. Since there is no assignment function which maps reference marker *a* to a witch, Hob's utterance is not true, though it is not false either. This prediction seems to accord with our untutored intuitions: our untutored intuitions dictate that Hob's utterance is not false because it has a false referential presupposition. (What is it then for a token τ relative to a DRS D̈ to be *false*? Once truth for an occurrence is adequately defined (See note 27), falsity for an occurrence could be defined as follows: A token τ of a sentence Σ uttered relative to DRS D which induces a DRS D' is *false* iff a token of the negation of Σ would be true relative to DRS D.)

It is important that according to the above provisional definition of truth for an occurrence, the *truth value* of an *occurrence* of a simple declarative sentence will almost always be identical to the *truth value* of the *proposition presented* by the occurrence (assuming that the occurrence has a truth value, and presents a proposition). For

example, *an occurrence* of 'Rover chased Spot' relative to DRS-1a' would be true if and only if there were an appropriate assignment function $f(\ )$ such that (i) $f(c)$ chased $f(b)$; $f(b)$ is a dog named 'Spot'; $f(c)$ is a dog named 'Rover'; and (ii) $f(b)$ is Spot; and $f(c)$ is Rover. But if there were such an assignment function, then the Russellian proposition *presented* by this occurrence, viz. that Rover chased Spot, would also be true. Conversely, if the Russellian proposition that Rover chased Spot is true, then there must be an appropriate assignment function $f(\ )$ such that $f(c)$ chased $f(b)$. Hence if it is assumed that Rover and Spot are dogs named 'Rover' and 'Spot,' respectively, it follows that an occurrence of 'Rover chased Spot' relative to DRS-1a' is true if and only if the proposition presented by this occurrence is true. Thus DRT predicts that, for occurrences of simple declarative sentences, differences between the truth values of occurrences and the truth values of the propositions presented by those occurrences will not, in normal situations, arise. As I will now begin to illustrate however, the truth values of occurrences of *attitude ascriptions* will often diverge from the truth values of the propositions these occurrences present.

How does this account of truth for occurrences apply in the case of attitude ascriptions? In other words, how does an occurrence of an attitude ascription affect the DRS relative to which it is uttered? The universe of a DRS for a discourse contains reference markers which allegedly represent the "entities" to which the participants of the discourse are existentially committed. For example, *a* is in the universe of DRS-1a because Hob and Nob are committed to there being a certain witch, and their utterances presuppose the existence of such a witch. Note, however, that in some cases an NP appearing within the that-clause of an ascription carries with it an existential presupposition, while in other cases such an NP carries no such presupposition. For example consider an utterance of

(3) Odile believes that Twain is a great author.

In most cases the utterer of (3) presupposes that 'Twain' has a referent. That is, because 'Twain' is a definite NP, the occurrence of 'Twain' must correspond with a reference marker which is linked to a reference marker already present in the universe of the DRS. In these cases the speaker will assent to an "exportation version" of his utterance. That is, if an occurrence of 'Twain' corresponds with a reference marker which is linked to a reference marker already in the DRS for the discourse, then the speaker would assent to an occurrence of 'There is somebody such that Odile believes that he is a great author'. (Note that the speaker might assent to this "exportation version" of his utterance, and yet not assent to a "substitution version" of his utterance relative to this DRS; i.e. he might assent to this "exportation version" of (3), and yet dissent from an occurrence of (4).) In other cases, however, the utterer of (3) does not presuppose that 'Twain' has a referent, and in these cases the speaker would *dissent from* the "exportation version" of his utterance. For example, imagine that Otho, who is skeptical concerning the existence of Twain, utters the following discourse:

(40) Odile believes that there is a guy called 'Twain', and that Twain is a great author.

In this case the speaker, Otho, does not presuppose that there is a guy called 'Twain', nor would he assent to an "exportation version" of his utterance. It is clear, however that in Otho's utterance of (40) 'Twain' is in some sense intended to be "coreferential" with the indefinite NP 'a guy called 'Twain''. Hence the reference markers corresponding to the two NPs must be linked to the same reference marker, yet there can be no such reference marker(s) in the DRS of Otho's discourse, as Otho does not presuppose that there is a guy called 'Twain'. How can this difference concerning the presuppositions carried by NPs appearing inside the that-clauses of occurrences of attitude ascriptions be represented within the framework of DRT?

270

In making attitude ascriptions, in speaking about how others believe, or fear, etc., the world to be, we endeavor to add to the context of our discourse propositions concerning what propositions the subjects of our ascriptions believe, fear, etc. Hence discourses concerning a person's beliefs can be thought of as involving a *sub-context* for that person: a *sub-context* for z in the context for x and y's conversation contains all and only those propositions mutually believed by x and y to be believed by z. Corresponding to contexts with sub-contexts there are DRSs with sub-DRSs. The difference with regard to the existential presuppositions carried by terms appearing inside that-clauses can be accounted for by appeal to DRSs with sub-DRSs. For example, suppose that Oscar's utterance of (3), the utterance for which 'Twain' carries existential presupposition, occurs relative to the following initial DRS:

DRS-3a

    Universe: *a, b*

    Conditions: 1. *a* is named 'Odile'
                    2. *b* is named 'Twain'
                    3. *b* is named 'Clemens'

How does Oscar's utterance of (3), where 'Twain' as it appears in this utterance carries an existential presupposition, affect DRS-3a? I suggest that the resulting DRS is

<u>DRS-3b</u>

Universe: *a, b,*

Conditions: 1. *a* is named 'Odile'
                 2. *b* is named 'Twain'
                 3. *b* is named 'Clemens'
                 4. *a* is in a belief state which is
                    characterized by the following sub-DRS:

        Universe: *c,*

        Conditions: 1. *c* is a great author.[28]

Internal Restrictions: 1. $c \leftrightarrow b$

External Restrictions: 1. $R(a) = $ Odile
                              2. $R(c) = $ Twain

A few points of clarification: First, the sub-DRS appears as part of condition 4 of the *primary* DRS, DRS-3b. Thus, since the occurrence of (3) relative to DRS-3a adds condition 4, if this occurrence is to be true, there must be an appropriate assignment function which, *inter alia*, satisfies this condition. But what is it for condition 4 to be satisfied under an assignment function? I.e. what is it for a sub-DRS to *characterize* a belief state under an assignment function $f(\ )$? A sub-DRS D of a primary DRS D' *characterizes* a belief state of a subject *x* under an assignment function $f(\ )$ which is appropriate for D' if and only if all the *propositions determined by the conditions of D under $f(\ )$* are elements of *x*'s belief set. What propositions are *determined by* a given sub-DRS under a given assignment function $f(\ )$? If all of the conditions of a given sub-DRS contain reference markers which are (i) linked to elements of the universe of the primary DRS, and (ii) linked to elements governed by external restrictions, then the sub-DRS determines the Russellian propositions which are obtained by combining the value of $R(\ )$ for each reference marker with the property designated by the appropriate condition. E.g. the sub-DRS of primary DRS-3b determines the Russellian proposition

that Twain is a great author. (Intuitively, the sub-DRS of DRS-3b determines this proposition because 'Twain' is assumed to have a referent and it does have a referent.) A sub-DRS also determines what I will call *fusion propositions*. A fusion proposition is obtained by forming "lambda properties" from conjunctions of conditions, and applying these properties to the entities represented by the reference markers which appear in the conditions. For example, suppose that the sub-DRS of DRS-3b also contained the following condition, 2. c is witty. What propositions would be determined by this sub-DRS? First, as described above, this sub-DRS would determine the propositions that Twain is a great author, and that Twain is witty. We now replace the terms referring to reference markers appearing in the conditions with variables, and conjoin the resulting open sentences. Thus in our example we obtain, 'x is a great author & x is witty'. We now form a lambda predicate, viz. '$\lambda x[x$ is a great author & x is witty]', and apply the property designated by this predicate to the entity, or entities, represented by the reference marker(s) which originally appeared in the conditions--reference marker a in our example. Thus a sub-DRS which contained conditions 1. a is a great author, and 2. a is witty, where a is an element of the universe of the primary DRS and R(a) = Twain, determines the following three propositions: that Twain is a great author, that Twain is witty, and that $\lambda x[x$ is a great author & x is witty](Twain). Thus such a sub-DRS would *characterize* a subject's belief state just in case the subject held the attitude of belief toward these three propositions.

What if the reference marker(s) appearing in a conditions are not linked to a reference marker in the universe of the primary DRS? Conditions containing such "free" reference markers determine existentially quantified Russellian propositions, and sets of conditions containing such "free" reference markers determine existentially quantified fusion propositions. For instance, in the above example if reference marker c were not linked to reference marker b, then the sub-DRS of DRS-3b would determine the Russellian proposition that there is somebody who is a great author. Moreover, if c were

273

not linked to b, and the sub-DRS of DRS-3b contained condition 2, then the sub-DRS would determine the fusion proposition that there is somebody who is a great author who is named 'Twain'.[29] (Note that a sub-DRS containing the conditions, 1. a is witty, and 2. b is not witty, do *not* determine the conjunctive Russellian proposition that Twain is witty *and* Twain is not witty, even if both a and b are governed by external restrictions which force them to be assigned to Twain.)

Second (continuing with the clarificatory remarks), DRS-3b contains an *Internal Restriction*, and nothing thus far has been said about these. Internal restrictions impose constraints upon assignment functions by specifying links between the reference markers in the universes of primary DRSs and reference makers in the universes of their sub-DRSs. An internal restriction represents that the participants of a discourse assume, or mutually believe, of two or more NP occurrences that they are to be *understood as* coreferrential (or codenoting). (We may *understand* occurrences of 'Santa' and 'St. Nick' *as* coreferring terms, even if we are in the know about Santa.) Thus the internal restriction in the above DRS represents that the participants of the discourse assume that the occurrence of 'Twain' appearing in Oscar's utterance of (3) is coreferential with past occurrences of 'Twain' and 'Clemens'. Roughly put, the internal restriction in DRS-3b represents that the participants assume that the person whom Odile is said to have a belief about is Twain, the same guy assumed by all participants to exist, and to be named both 'Twain' and 'Clemens'. Note that Oscar's utterance contains appearances of two successfully referring definite direct reference NPs, viz. 'Odile' and 'Twain'. Thus two external restrictions are added, and each NP occurrence corresponds with a reference marker which is linked to an already present reference marker. As is indicated by condition 4, the occurrence of the definite NP 'Odile' corresponds with and is linked to already present reference marker c: since this occurrence of 'Odile' appears outside of any that-clause, it does not introduce a new reference marker into the DRS. The occurrence of 'Twain', on the other hand, appears inside a that-clause, and consequently it

274

corresponds with reference marker c, which is introduced into the universe of the sub-DRS. As is required by the Familiarity Constraint, this new reference marker is then linked via the internal restriction to reference marker b.

Putting the above points together, Oscar's utterance of (3) relative to DRS-3a is true just in case there is an appropriate assignment function $f( )$ such that $f(c)=f(b)=R(c)=$Twain; $f(b)$ is named 'Twain and 'Clemens'; and the proposition that Twain is a great author is an element of Odile's belief set. (This is the only proposition determined by the sub-DRS of DRS-3b under assignment function $f( )$.)

We have been assuming that in Oscar's utterance of (3) relative to DRS-3a the appearance of 'Twain' carries an existential presupposition; we have been assuming that Oscar presupposes that this appearance of 'Twain' has a referent. That this appearance of 'Twain' carries an existential presupposition is reflected in DRS-3b: the internal restriction of DRS-3b links b, which is an element of the universe of DRS-3b, with c, the reference marker introduced by Oscar's utterance of 'Twain'. How can an utterance of

(40) Odile believes that there is a guy called 'Twain', and that Twain is a great author.

in which the appearance of 'Twain' does not carry an existential presupposition, be represented in the framework of DRT? Note that Otho's utterance of (40) contains an appearance of the indefinite NP, 'a guy'. Hence the reference marker that corresponds with this NP cannot be linked to a reference marker already present anywhere in the initial DRS. Let us assume therefore that the initial DRS, the DRS relative to which Otho utters (40), is as follows:

275

<u>DRS-4a</u>

Universe: *a*

Conditions: 1. *a* is named 'Odile'

Let us take the two conjuncts of (40) in turn. I suggest that the DRS resulting from Otho's utterance of 'Odile believes that there is a guy called 'Twain',' relative to DRS-4a induces

<u>DRS-4b</u>

Universe: *a*

Conditions: 1. *a* is named 'Odile'
            2. *a* is in a belief state which is
               characterized by the following sub-DRS:

               Universe: *b*,

               Conditions 1. *b* is called 'Twain'
                         2. *b* is a guy

External Restrictions: 1. R(*a*) = Odile

The indefinite NP 'a guy', since it appears inside a that-clause, introduces reference marker *b* into the sub-DRS. (Note that two conditions are added to the sub-DRS by Otho's utterance: the first condition is obtained by replacing the introducing NP with the corresponding reference marker, the other is obtained by predicating information conveyed by the NP occurrence itself to the introduced reference marker. That occurrences of NPs themselves introduce conditions into sub-DRSs will be important in what follows.) As usual, the reference marker corresponding with the successfully referring directly referential NP 'Odile' is governed by an external restriction.

276

Now consider the second conjunct of (40): Otho's utterance of 'that Twain is a great author', relative to DRS-4b induces

<u>DRS-4c</u>

Universe: a

Conditions: 1. a is named 'Odile'
2. a is in a belief state which is
characterized by the following sub-DRS:

Universe: b,

Conditions: 1. b is called 'Twain'
2. b is a guy
3. b is a great author

External Restrictions: 1. $R(a)$ = Odile
2. $R(b)$ = Twain

Note that the truth of Otho's utterance does not require that b be assigned to an actual entity; b is an element of the sub-DRS, and there is no internal restriction linking b to some element of the universe of the primary DRS. (This indicates that Oscar does not assume that 'Twain' has a referent, even though it does.) Thus Otho's utterance could be true even if there were no appropriate assignment function $f(\ )$ such that the conditions of the sub-DRS were satisfied under $f(\ )$. More specifically, Otho's utterance of (4) relative to initial DRS-4a is true just in case there is some appropriate assignment function $f(\ )$ such that, $f(a)$ = Odile, and $f(b)$=Twain; $f(a)$ is named 'Odile'; and Odile holds the attitude of belief toward the following Russellian propositions: that someone named 'Twain', that someone is a guy, that someone is a great author, and that someone who is a guy named 'Twain' is a great author. (Note that the sub-DRSs determine existential propositions, as opposed to singular propositions, because b is not linked to a reference

marker in the universe of DRS-4c, and an appropriate assignment function for DRS-4c is undefined for argument *b*.)

We are almost ready to illustrate how DRT can be invoked to distinguish the truth conditions of occurrences of (3) and (4), but a few more pieces of machinery must be put into place. Consider again DRS-3b, the DRS which was induced from Oscar's utterance of (3) relative to DRS-3a:

<u>DRS-3b</u>

    Universe: *a, b*,

    Conditions: 1. *a* is named 'Odile'
             2. *b* is named 'Twain'
             3. *b* is named 'Clemens'
             4. *a* is in a belief state which is
                characterized by the following sub-DRS:

        Universe: *c*,

        Conditions: 1. *c* is a great author.

    Internal Restrictions: $c \leftrightarrow b$

    External Restrictions: 1. R(*a*) = Odile
                       2. R(*c*) = Twain

If this were the DRS which resulted from an utterance of (3), it also could be the DRS which resulted from an utterance (4). The only difference between (3) and (4) is that (3) has 'Twain' where (4) has 'Clemens', but the fact that (3) has 'Twain' instead of 'Clemens' appearing in its that-clause is not reflected anywhere in the above DRS. Consequently, Oscar's utterance would have had the same truth conditions had he uttered (4) instead of (3). In other words, if Oscar's utterance of (3) relative DRS-3a induces DRS-3b, then the context which DRS-3a *partially characterizes* is an *insensitive* context. Consequently, as it is often put, DRS-3b represents a *de re* interpretation of Oscar's utterance. What is

278

needed in order to distinguish the truth conditions of occurrences of (3) and (4) is some way of representing a *de dicto* interpretation of Oscar's utterance within the framework of DRT.[30]

Consider an utterance of

(41) Odile believes that the witty guy is a great author.

relative to DRS-5a:

<u>DRS-5a</u>

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is a witty guy

How can the *de dicto* interpretation of this utterance relative to DRS-5a be represented? That is, how might an utterance of (41) relative to DRS-5a determine a DRS which has distinct truth conditions from the DRS determined by an utterance of

(42) Odile believes that the bearded guy is a great author.

relative to DRS-5a, even though the occurrences of 'the witty guy', and 'the bearded guy', denote the same individual? I suggest that the following represents a *de dicto* interpretation of the utterance of (41) relative to DRS-5a:

<u>DRS-5b</u>

Universe: *a, b,*

Conditions: 1. *a* is named 'Odile'
2. *b* is a witty guy
3. *a* is in a belief state which is
characterized by the following sub-DRS:

Universe: *c,*

Conditions: 1. *c* is a witty guy
2. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: R(*a*) = Odile

Because the definite NP 'the witty guy' appears inside a that-clause, it corresponds with

the reference marker *c* which is introduced into the universe of the sub-DRS. And

because the NP is definite, *c* must be linked, via the internal restriction, to an already

present reference marker. If DRS-5b is the DRS brought about by the occurrence of (41)

relative to DRS-5a, then the truth of this occurrence requires that there be an appropriate

assignment function $f(\ )$ such that $f(b)$ is a witty guy and that Odile have the fusion

proposition that $f(b)$ is a witty guy who is a great author in her belief set. Contrast the

above with DRS-5b', which represents the *de dicto* reading of an utterance of (42)

relative to DRS-5a:

<u>DRS-5b'</u>

Universe: *a, b,*

Conditions: 1. *a* is named 'Odile'
                2. *b* is a bearded guy
                3. *a* is in a belief state which is
                   characterized by the following sub-DRS:

Universe: *c,*

Conditions: 1. *c* is a bearded guy
                   2. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: R(*a*) = Odile

The truth of the utterance of (42) relative to DRS-5a thus requires that there be an appropriate assignment function $f(\ )$ such that $f(b)$ is a bearded guy and that Odile have the fusion proposition that $f(b)$ is a bearded guy who is a great author in her belief set. Thus if DRS-5b and DRS-5b' are brought about by the utterances of (41) and (42) respectively, then these occurrences have distinct truth conditions. In other words, DRS-5b and DRS-5b' represent *de dicto* interpretations of utterances of (41) and (42). The *de dicto* interpretation of the utterance of (41) is represented by placing the condition, 1. *c* is a witty guy, which is specified by the NP 'the witty guy', into the sub-DRS. Similarly, the *de dicto* interpretation of the occurrence of (42) is obtained by placing the condition, 1. *c* is a bearded guy, which is specified by the NP 'the bearded guy', into the sub-DRS. Note that these NP specified conditions might instead have been placed in the primary DRSs, in which case the resulting DRSs would have represented *de re* interpretations.

Before applying this machinery to occurrences of (3) and (4), I would like to highlight an advantage of DRT which has been brought out by the discussion thus far. It is usually assumed that there are two kinds of belief ascriptions: *de re* ascriptions, and *de*

dicto ascriptions. (Sometimes it is suggested that there is a third category of *"de se"* attitude ascriptions. E.g. "Heimson believes that he [himself] is Hume." An example of a *de se* ascription is discussed in Appendix B.[31]) As was brought out by the above discussion, however, there are actually three kinds of belief ascriptions. For any ascription of the form N believes that $\Sigma(\tau)$, where τ is a singular definite NP and Σ is a sentential function, there are three kinds of interpretation:

(i) The ascription may be interpreted *de re*, where the occurrence of τ carries an existential presupposition.

(ii) The ascription may be interpreted *de dicto*, where the occurrence of τ carries an existential presupposition.

(iii) The ascription may be interpreted *de dicto*, where the occurrence of τ does not carry an existential presupposition.

(Interpretations of type (ii) are often overlooked. That there are cases of type (ii) is evidenced by the existence of situations in which a speaker assents to a token of 'Odile believes that Twain is a great author' and also a token of 'There is somebody whom Odile believes to be a great author' yet dissents from a token of 'Odile believes that Clemens is a great author.' Also, note that there could not be a *de re* interpretation where τ carried no existential presupposition.) The DRT machinery thus far sketched accounts for this three-way distinction. Consider an utterance of

(41) Odile believes that the witty guy is a great author.

relative to the following initial DRS:

<u>DRS-5a</u>

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
　　　　　　2. *b* is a witty guy

There are three ways of interpreting an occurrence of (41) relative to DRS-5a.

Interpretation (i) is represented by the following DRS:

<u>DRS-5bi</u>

Universe: *a, b,*

Conditions: 1. *a* is named 'Odile'
　　　　　　2. *b* is a witty guy
　　　　　　3. *a* is in a belief state which is
　　　　　　　　characterized by the following sub-DRS:

　　　　　　　　Universe: *c,*

　　　　　　　　Conditions: 1. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: R(*a*) = Odile

The second interpretation of an utterance of (41) relative to DRS-5a is represented by the

following DRS:

<u>DRS-5bii</u>

Universe: *a, b,*

Conditions: 1. *a* is named 'Odile'
            2. *b* is a witty guy
            3. *a* is in a belief state which is
              characterized by the following sub-DRS:

        Universe: *c*

        Conditions: 1. *c* is a great author
                        2. *c* is a witty guy

Internal Restrictions: $c \leftrightarrow b$

External Anchors: $R(a) = $ Odile

And finally the third interpretation is represented by the following DRS:

<u>DRS-5biii</u>

Universe: *a,*

Conditions: 1. *a* is named 'Odile'
            2. *a* is in a belief state which is
              characterized by the following sub-DRS:

        Universe: *c*

        Conditions: 1. *c* is a great author
                    2. *c* is a witty guy

External Restrictions: $R(a) = $ Odile

(Note that interpretation (iii) is not available if (41) is uttered relative to DRS-5a.

Interpretation (iii) is available only if (41) is uttered relative to a DRS which already

contains a sub-DRS, and the occurrence of the definite NP 'the witty guy' must

correspond with a reference marker which is linked to a member of the universe of this

sub-DRS. For, example, DRS-5biii could be induced from an initial DRS which contained a sub-DRS with r in its universe, and r *is witty* among its conditions.) Thus DRT can account for all three interpretations of an occurrence of an attitude ascription. I know of no other theory of attitude ascriptions which is able to do this.

How can this machinery be utilized in order to represent *de dicto* interpretations of occurrences of (3) and (4)? The basic idea is that in order to obtain *de dicto* interpretations of utterances of these sentences, the NPs 'Twain' and 'Clemens' must also specify conditions which are placed in the sub-DRSs of the DRSs which are induced by these utterances. Just as an utterance of

(41) Odile believes that the witty guy is a great author.

can be interpreted in such a way that it is true only if Odile believes for some individual i that i is a witty guy who is a great author, so an utterance of

(3) Odile believes that Twain is a great author.

can be interpreted in such a way that it is true only if Odile believes that Twain is named 'Twain' and is a great author. An occurrence of (3), in a sensitive context, *presents* the Russellian proposition that Odile believes that Twain (the man) is a great author, but it also *shows* (to borrow Wittgenstein's term) and thereby *communicates* the Russellian proposition that Odile believes that Twain is named 'Twain'. Similarly, an occurrence of

(4) Odile believes that Clemens is a great author.

presents this same Russellian proposition, but it *shows* and thereby *communicates* the Russellian proposition that Odile believes that Twain is named 'Clemens'. Thus what the

machinery of DRT does, in essence, is to take Salmon's insights concerning how the names appearing in ascriptions affect our *intuitions* concerning truth conditions, and to make these insights applicable to the actual truth conditions of occurrences. Again, I maintain that Salmon is roughly correct in explaining why we have the intuitions we do, but he is incorrect in thinking that these intuitions are not veridical; his mistake is his tacit endorsement of the Principle of Propositional Truth.

At last the stage is set and the above insights can be applied within the framework of DRT in order to distinguish the truth conditions of occurrences of (3) and (4). Let us suppose that the initial DRS for an utterance of (3) is

DRS-6a

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'

Assuming that the goals of the discourse are such that the context which is partially characterized by DRS-6a is *sensitive*, the utterance of (3) relative to DRS-6a induces

<u>DRS-6b</u>

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'
4. *a* is in a belief state which is characterized by the following sub-DRS:

Universe: *c*

Conditions: 1. *c* is named 'Twain'
2. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: 1. R(a) = Odile
2. R(c) = Twain

The utterance of (3) relative to DRS-6a is thus true if and only if there is some assignment function $f(\ )$ such that $f(\ )$ obeys all the external and internal restrictions of DRS-6b, and conditions 1-4 are satisfied under $f(\ )$. Thus the occurrence of (3) is true if and only if there is an appropriate assignment function $f(\ )$ such that, $f(a)$ = Odile and $f(b) = f(c)$ = Twain; $f(a)$ is named 'Odile', and $f(b)$ is named 'Twain' and 'Clemens'; and finally $f(a)$ holds the attitude of belief toward the Russellian propositions determined under $f(\ )$ by the conditions in the sub-DRS of condition 4; i.e. Odile must hold the attitude of belief toward the Russellian propositions that Twain is named 'Twain', that Twain is a great author, and that Twain is a great author named 'Twain'. (This last proposition is more formally presented by '$\lambda x[x$ is a great author & x is named 'Twain'](Twain)'.)

Now let us determine the truth conditions of an utterance of (4) relative to DRS-6a. Assuming that the context is a sensitive context, an utterance of (4) relative to DRS-6a induces the DRS

287

<u>DRS-6b'</u>

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'
4. a is in a belief state which is
characterized by the following sub-DRS:

Universe: *c,*

Conditions: 1. *c* is named 'Clemens'
2. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: 1. R(*a*) = Odile
2. R(*c*) = Twain

The utterance of (4) relative to DRS-6a, similar to the occurrence of (3) considered

above, is true if and only if there is some appropriate assignment function $f( )$ such that,

$f( )$ obeys all the external and internal restrictions of DRS-6b', and conditions 1-4 are

satisfied under $f( )$. Thus the occurrence of (4) is true if and only if there is an

appropriate assignment function $f( )$ such that, $f(a)$ = Odile and $f(b) = f(c)$ = Twain;

$f(a)$ is named 'Odile', $f(b)$ is named 'Twain' and 'Clemens'; and finally $f(a)$ holds the

attitude of belief toward the Russellian propositions determined under $f( )$ by the

conditions in the sub-DRS of condition 4; i.e. Odile must hold the attitude of belief

toward the Russellian propositions that Twain is named 'Clemens', that Twain is a great

author, and that Twain is a great author named 'Clemens'. (Again, this last proposition is

more formally presented by, '$\lambda x[x$ is a great author & x is named 'Clemens'](Twain)'.)

Thus DRT illustrates how, in sensitive contexts, the truth conditions of occurrences of (3)

and (4) differ, even though these occurrences present the very same Russellian proposition.

What about the truth conditions of negated ascriptions? If a token of

(4n) Odile does not believe that Clemens is a great author.

is uttered relative to DRS-6a, the following DRS is induced:

DRS-6bn'

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'
4. *a* is *not* in a belief state which is characterized by the following sub-DRS:

Universe: *c,*

Conditions: 1. *c* is named 'Clemens'
2. *c* is a great author

Internal Restrictions: $c \leftrightarrow b$

External Restrictions: 1. $R(a)$ = Odile
2. $R(c)$ = Twain

The utterance of (4n) is true relative to DRS-6a if and only if there is some appropriate assignment function $f(\ )$ such that $f(\ )$ obeys all the external and internal restrictions of DRS-6bn', and conditions 1-4 are satisfied under $f(\ )$. Thus the occurrence of (4n) is true if and only if there is an appropriate assignment function $f(\ )$ such that, $f(a)$ = Odile and $f(b) = f(c)$ = Twain; f(a) is named 'Odile', $f(b)$ is named 'Twain' and 'Clemens'; and finally, $f(a)$ does *not* hold the attitude of belief toward all the Russellian propositions determined under $f(\ )$ by the conditions in the sub-DRS of condition 4; i.e. Odile does

not hold the attitude of belief toward at least one of the following propositions: that Twain is named 'Clemens', that Twain is a great author, and that Twain is a great author named 'Clemens'.

Have I succeeded in motivating option (iii)? Have I succeeded in showing that it is plausible to maintain that, in keeping with the Russellian desiderata, occurrences of (3) and (4) present the same Russellian proposition, yet even so our untutored intuitions are correct in dictating that occurrences of (3) and (4) may have distinct truth conditions? I concede that the above sketch of DRT is insufficient in that it does not address many questions and concerns which need to be addressed. Three of the most pressing concerns are as follows: First, an adequate account of DRT would have to be formalized with much greater precision. In particular, instead of relying on intuitions, a set of formal rules for constructing DRSs from discourses would have to specified. (Kamp calls the set of rules for constructing DRSs from discourses the "construction algorithm."[32]) Furthermore, the above provisional definition of truth for an occurrence must be modified so that it is intuitively adequate; the provisional account of truth for an occurrence must be amended so that the truth value of an occurrence does not depend upon every preceding occurrence in the discourse .

Second, the relationship between DRT and conventional truth-theoretic semantic theories must be clarified. I have assumed here that DRT does not *supplant* traditional truth-theoretic semantics, but rather *complements* such theories. That is, I have assumed that something roughly along the lines of the Naive Russellian Theory--minus the Principle of Propositional Truth--will give a correct account of the propositions presented by occurrences; DRT was brought in to show how such a semantic theory can be made to accord with our untutored intuitions. It may turn out, however, that a comprehensive theory of discourse representation will itself be able to explain the relevant phenomena, without recourse to anything much resembling the Naive Russellian Theory (minus the Principle of Propositional Truth).[33] I have no argument to the effect that DRT will not

in the end supplant traditional truth theoretic semantics, but I do not see the prospect of

DRT supplanting traditional truth theoretic approaches to semantics as a threat to the

central thesis of this dissertation. The central thesis of this dissertation is that a theory

which endorses both the Principle of Direct Reference and the Principle Semantic

Innocence can be defended from the problems and arguments posed by the phenomenon

of opacity. Therefore, so long as the supplanting discourse representation theory respects

principles at least similar to the principles of direct reference and semantic innocence, the

central thesis of this dissertation stands.

Third, and I think most importantly, a more precise statement of what discourse

representation structures *are* is required. What exactly are reference markers? What

exactly is the relationship between discourse representation structures, and what goes on

when actual people engage in actual discourses? My purpose in this section has been to

illustrate how DRT can be utilized to defend Russellian theories from the semantic

arguments from opacity, and thus I have not addressed these questions. An adequate

defense of DRT, however, would be required to provide answers to these difficult

questions.[34]

Despite these deficiencies however, I maintain that the above sketch of DRT is

sufficient to justify the claim that option (iii) is superior to either option (ii), or option (i).

That is, the appropriate response to the semantic arguments from opacity is not to invent

a complicated way of distinguishing the propositions presented by attitude ascriptions,

nor is it to deny the veracity of our untutored intuitions concerning occurrences of

attitude ascriptions. Rather the appropriate response to the semantic arguments from

opacity is to deny the Principle of Propositional Truth, and to provide an account of truth

for an occurrence which accords with our untutored intuitions.

Is DRT, as sketched above, something to which a Russellian theorist may avail

himself? There are, to be sure, similarities between reference markers and Fregean

senses, and I have argued in previous chapters that nothing can play the role that senses

were posited to play. Is DRT, as sketched above, compatible with my arguments against the Fregean strategy? First, it should be noted that DRT is wholly compatible with the Russellian desiderata, viz. the principles of direct reference, and semantic innocence. DRT is wholly compatible with the claim that the objects, or contents, of propositional attitudes are Russellian propositions, and Russellian propositions are the semantic values of occurrences of declarative sentences. And DRT is wholly compatible with the claim that a declarative sentence presents the same Russellian proposition relative to a context regardless of whether the sentence appears inside or outside of a that-clause. Furthermore, unlike theories which take option (i) (i.e. maintaining the Principle of Propositional Truth, and inventing a way of distinguishing the propositions presented by ascriptions such as (3) and (4)), option (iii) as realized by DRT does not require ordinary speakers to refer to, or have the ability to specify, senses, sentences of mentalese, cognitive particulars, or any other kind of esoteric mental entities. Thus option (iii) as realized by DRT is not squeezed between the Scylla and Charybdis of the Fregean strategy: The mental goings on which account for a subject's behavior play no part in determining what is presented, or even communicated, by an occurrence of an attitude ascription about that subject. In making attitude ascriptions speakers are not attempting to refer to, nor even to represent, the mental goings on of the subject. Rather, in uttering an attitude ascription a speaker is attempting to describe the world as the subject believes, desires, etc. it to be. Of course in so doing the speaker, if his utterance is true, thereby *characterizes* the mental state of the subject, but characterizing a mental state does not require referring to, or representing, components or features of that mental state.

4.4 A DRT Analysis of "Folk" Explanations of Behavior.

In this section I will sketch a strategy for utilizing the machinery of DRT to account for the truth conditions of occurrences of "folk" explanations of behavior. I must warn the reader that my remarks on this complicated issue leave many questions unanswered, and thus are unfortunately rather programmatic in nature.

Recall the problem that occurrences of "folk" explanations of behavior pose for the Naive Russellian Theory: Our untutored intuitions dictate that, in sensitive contexts, occurrences of

(10a) Odile desires that Twain come to the party

and

(10b) she believes that if wine is served, then Twain will come to the party

and *that's why*

(10c) she is going to serve wine.

are true, while occurrences of

(10a) Odile desires that Twain come to the party

and

(10b') she believes that if wine is served, then Clemens will come to the party

and *that's why*

(10c) she is going to serve wine.

are false. The difficulty presented by such folk explanations is not merely that of accounting for why the occurrences of the sentential constituents (10b) and (10b') can

have distinct truth conditions--that is just the problem posed by the semantic arguments from opacity. Rather the problem presented by such folk explanations is to account for why what is conveyed by occurrences of (10a) and (10b) might serve as *an explanation* of what is conveyed by (10c), while what is conveyed by occurrences of (10a) and (10b') might not serve as *an explanation* of what is conveyed by (10c). The problem is not that of accounting for the truth conditions of the sentential constituents of occurrences of (10a-b-c) and (10a-b'-c), rather the problem is that of accounting for the truth conditions of occurrences of (10a-b-c) and (10a-b'-c) as whole sentences.

The Naive Russellian Theory is unable to distinguish the truth conditions of common belief\desire explanations such as (10a-b-c) and (10a-b'-c) because it maintains that the that-clauses of occurrences of (10b) and (10b') have the same Russellian proposition as semantic value. Therefore, since occurrences of (10a-b-c) and (10a-b'-c) present the same Russellian proposition and the Naive Russellian Theory endorses the Principle of Propositional Truth, it follows that the Naive Russellian Theory cannot distinguish the truth conditions of occurrences of (10a-b-c) and (10a-b'-c).

The machinery introduced in section 4.3, however, provides a means of individuating the truth conditions of occurrences of (10a-b-c) and (10a-b'-c). Let us assume that the DRS relative to which (10a) is uttered is

DRS-7a

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'

An utterance of (10a) relative to DRS-7a induces

<u>DRS-7b</u>

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
          2. *b* is named 'Twain'
          3. *b* is named 'Clemens'
          4. *a* is in a *mental state* which characterized
             by the following sub-DRS:

             Universe: *c*

             B-conditions: 1. *c* is named 'Twain'

             D-conditions: 1. *c* comes to the party

Internal Restrictions: 1. c ↔ b

External Restrictions: 1. R(*a*) = Odile
                      2. R(*c*) = Twain

Reference marker *a* corresponds with the occurrence of the directly referential NP 'Odile', while reference marker *c* corresponds with the occurrence of the directly referential NP 'Twain'. As these directly referential NPs successfully refer, external restrictions 1 and 2 are added to DRS-7a. Since these occurrences are occurrences of definite NPs, internal restriction 1 is added to DRS-7a, linking the introduced reference marker *c* to already present reference marker *b*. Note that the sub-DRS of condition 4 allegedly characterizes the *mental state* of Odile, and thus contains both "B-conditions" (Belief-conditions) and "D-conditions" (Desire-conditions). This new wrinkle in DRS formation is required to account for the fact that *de dicto* desire ascriptions (and *de dicto* regret ascriptions, etc.) often communicate propositions concerning what the subject believes. For example, in a sensitive context an utterance of

(43) Odile desires that the witty guy come to the party.

295

communicates that Odile *believes* that there is a guy who is witty. Similarly, in a

sensitive context an utterance of

(10a) Odile desires that Twain go to the party.

communicates that Odile *believes* that Twain is named 'Twain'.

An utterance of (10b) relative to DRS-7b induces

DRS-7c

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
          2. *b* is named 'Twain'
          3. *b* is named 'Clemens'
          4. *a* is in a mental state which is
             characterized by the following sub-DRS:

            Universe: *c*

            B-conditions: 1. *c* is named 'Twain'
                          2. If wine is served, then
                             c comes to the party

            D-conditions: 1. *c* comes to the party

Internal Restrictions: 1. *c* ↔ *b*

External Restrictions: 1. R(*a*) = Odile
                       2. R(*c*) = Twain

The occurrence of 'Odile' in (10b) corresponds with reference marker *a*, while the

occurrences of 'Twain' in (10b) corresponds with reference marker *c*. And internal

restriction 1 (still) links the reference marker *c* with the appropriate already present

reference marker *b*.

Finally, an utterance of (10c) relative to DRS-7c induces

296

DRS-7d

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
              2. *b* is named 'Twain'
              3. *b* is named 'Clemens'
              4. *a* is in a mental state which characterized
                 by the following sub-DRS:

      Universe: *c,*

      B-conditions: 1. *c* is named 'Twain'
                        2. If wine is served, then
                           c comes to the party

      D-conditions: 1. *c* comes to the party

      5. *a* is going to serve wine and this is
         explained by the fact that *a* is in a
         mental state characterized by condition 4.

Internal Restrictions: 1. $c \leftrightarrow b$

External Restrictions: 1. R(*a*) = Odile
                            2. R(*c*) = Twain

The occurrence of 'Odile' appearing in (10c) corresponds with reference marker *a*, which is already present in the DRS, and is already governed by an external restriction. Thus an assertion of (10a-b-c) in a sensitive context is true if and only if there is an appropriate assignment function which satisfies DRS-7d. The conditions of DRS-7d are all familiar, with the exception of condition 5, the only condition added to DRS-7c by the utterance of (10c). What is it for an appropriate assignment function to satisfy condition 5?

First, condition 5 is satisfied by an appropriate assignment function $f(\ )$ only if $f(a)$ is going to serve wine. If there is no such assignment function, the entire occurrence of (10a-b-c) is not true. But more than this is required for the satisfaction of condition 5.

Intuitively, condition 5 is satisfied only if the utterances of (10a) and (10b) offer an explanation of Odile's serving wine. But what is for the utterances of (10a) and (10b) to offer such an explanation?

I suggest that the utterances of (10a) and (10b) offer an explanation of Odile's serving wine if and only if the conditions of the sub-DRSs added by these utterances instantiate the appropriate structure. More precisely, the conditions of the sub-DRSs added by these utterances must instantiate the same basic structure as is displayed by the following familiar belief/desire explanation schemata:

(a) X desires C
(b) X believes (E occurs $\rightarrow$ C is realized)
Therefore,
(c) X acts so as to make E occur.

That is, I suggest that an occurrence of (10a-b-c) is true only if the structure displayed by the above schemata is instantiated *at level of discourse representation*. The reason that an occurrence of (10a-b'-c) might be false relative to a context in which an occurrence of (10a-b-c) is intuitively true is that the DRS induced by the occurrence of (10a-b-c) has the appropriate structure, while the DRS induced by the occurrence of (10a-b'-c) does not have the appropriate structure. More precisely, the conditions added by the occurrence of (10a-b-c) instantiate the structure displayed by the above schemata, while the conditions added by the occurrence of (10a-b'-c) do not instantiate this structure.

How is the structure of the above belief\desire explanation schemata instantiated in DRS-7d? Condition 4 of DRS-7d is satisfied only if there is an assignment function $f^*(\ )$ such that

(a') $f^*(a)$ believes that if wine is served, then $f^*(c)$ comes to the party.

and

$$(\text{b'}) \quad f^*(a) \text{ desires that } f^*(c) \text{ comes to the party.}$$

And condition 5 of DRS-7d is satisfied by $f^*(\ )$ only if

$$(\text{c'}) \quad f^*(a) \text{ is going to serve wine.}$$

I maintain that if $f^*(a)$ is in a mental state characterized by condition 4, then the second conjunct of condition 5 is satisfied by $f^*(\ )$ because (a')-(b')-(c') have the appropriate structure; i.e. because (a')-(b')-(c') can serve as an instance of the belief/desire explanation schemata:

(a) X desires C
(b) X believes (E occurs $\rightarrow$ C is realized)
Therefore,
(c) X acts so as to make E occur.

(Roughly, '$f^*(a)$' replaces schematic letter 'X'; ' that wine is served' replaces the schematic letter 'E'; and 'that $f^*(c)$ comes to the party' replaces schematic letter 'C'.) Thus condition 5 of DRS-7d is satisfied by an appropriate assignment function $f(\ )$ just in case (i) the first conjunct of condition 5 is satisfied by $f(\ )$, and (ii) the *sub-conditions* of condition 4, together with the first conjunct of condition 5, instantiate the structure displayed in the belief/desire schemata. The utterance of (10a-b-c) is true because, *inter alia*, the DRS induced by this utterance has the structure displayed by the familiar belief/desire explanation schemata.

What about an utterance of (10a-b'-c) relative to DRS-7a? I maintain that this utterance is false because the DRS induced by this utterance does not instantiate the

structure displayed by the familiar belief/desire explanation schemata. The only difference between (10a-b-c) and (10a-b'-c) concerns (10b) and (10b'). Thus an utterance of (10a-b'-c) relative to DRS-6a will induce a different DRS:

DRS-7d'

Universe: *a, b*

Conditions: 1. *a* is named 'Odile'
2. *b* is named 'Twain'
3. *b* is named 'Clemens'
4. *a* is in a mental state which characterized by the following sub-DRS:

Universe: *c, d*

B-conditions: 1. *c* is named 'Twain'
2. *d* is named 'Clemens'
3. If wine is served, then *d* comes to the party

D-conditions: 1. *c* comes to the party

5. *a* is going to serve wine and this is explained by the fact that *a* is in a mental state characterized by condition 4.

Internal Restrictions: 1. *c* ↔ *b*
2. *d* ↔ *b*

External Restrictions: 1. R(*a*) = Odile
2. R(*c*) = Twain
3. R(*d*) = Twain

The differences between DRS-7d and DRS-7d' are a result of the differences between

(10b) she believes that if wine is served, then **Twain** will come to the party

300

and

(10b') she believes that if wine is served, then **Clemens** will come to the party.

The occurrence of 'Twain' in the occurrence of (10b) corresponds with reference marker c, which is already present in the universe of the sub-DRS. Yet the occurrence of 'Clemens' in the occurrence of (10b') introduces a new reference marker, viz. d, into the universe of the sub-DRS. This indicates that as Odile believes the world to be, there may be two individuals, one named 'Twain' and the other named 'Clemens'. Note however, that the speaker presupposes that the occurrences of 'Twain' and 'Clemens' are coreferential, as is indicated by the internal restrictions of DRS-7d'. (Internal restriction 2 ensures that the Familiarity Constraint is satisfied with regard to the occurrence of 'Clemens'.)

Under what conditions is condition 5 of DRS-7d' satisfied by an appropriate assignment function? Condition 5 of DRS-7d' is satisfied by an appropriate assignment function $f^*( )$ just in case (i) the first conjunct is satisfied by $f^*( )$, and (ii) DRS-7d' has the appropriate structure; i.e. just in case the *sub-conditions* of condition 4, together with the first conjunct of condition 5, instantiate the structure displayed by the belief/desire schemata. Does DRS-7d' instantiate the appropriate structure? Condition 4 of DRS-7d' is satisfied only if there is an assignment function $f^*( )$ such that

(a') $f^*(a)$ believes that if wine is served, then $f^*(c)$ comes to the party.

and

(b") $f^*(a)$ desires that $f^*(d)$ comes to the party.

And the first conjunct of condition 5 of DRS-7d' is satisfied by $f*(\ )$ only if

(c') $f*(a)$ is going to serve wine.

Unlike sentences (a')-(b')-(c'), the sentences (a')-(b")-(c') do not serve as an instance of the belief\desire schemata. Consequently the second conjunct of condition 5 of DRS-7d' cannot be satisfied by an appropriate assignment function, and thus the occurrence of (10a-b'-c), relative to DRS-7a, is not true.

Thus DRT offers a means of accounting for the differing truth conditions of occurrences of (10a-b-c) and (10a-b'-c). The principal idea motivating the above account is that occurrences of "folk" explanations of behavior do not require for their truth that the *presented propositions* instantiate an appropriate explanatory structure, rather occurrences of such explanations require for their truth that an appropriate explanatory structure be instantiated *at the level of discourse representation*. Note, however, that the account of "folk" explanations sketched above does *not* require that occurrences of "folk" explanations refer to, or present, DRSs. Thus the above sketched account of the truth conditions of "folk" explanations is wholly compatible with the principles of direct reference and semantic innocence.

Chapter 4 Notes

[1] Stalnaker 1990, p. 131.

[2] Burge's model of self-knowledge is presented in Burge 1988. Similar proposals are sketched by Davidson 1987, and in Davidson 1984, and also in Stalnaker 1990. Burge's approach is criticized, wholly unsuccessfully in my view, in Boghossian 1989.

[3] Burge 1988, p. 654.

[4] Burge 1988, p. 658.

[5] Burge 1988, p. 658.

[6] Burge 1988, p. 659.

[7] Davidson 1988.

[8] Stalnaker 1987a, p. 15. A very similar dispositional analysis of propositional attitudes is presented and defended in Ruth Barcan Marcus 1990.

[9] Stalnaker 1987a, p. 22.

[10] The account of the content of a mental state stated here is a rough synopsis of the views Dretske proffers in Dretske 1981 (especially part III), and in Dretske 1988 (especially part III).

[11] If the above question can be answered in non intentional terms, then Dretske's analysis succeeds in solving the "problem of intentionality," the problem of explaining how a physical entity--such as an instance of a mental state--can be *about* something. I remain neutral as to whether or not the question can be answered in non intentional terms, but I think that, regardless of the success or failure of a Dretske-style theory on this score, the analysis of mental content in terms of *indication* and *function* is edifying.

[12] It is not clear whether a similar "self-verifying" phenomenon occurs with regard to propositional attitudes other than belief. The attitude of belief is unique in that one can indicate one's belief in what is expressed by an occurrence by *assenting to* the occurrence. There is no similar way in which one can indicate, for instance, that one holds the attitude of *hope* toward what is expressed by an occurrence. If one wants to know whether or not a subject holds the attitude of hope toward what is expressed by Σ, it seems that one must ask the subject, ⌜Do you hope that Σ?⌝ If the subject responds affirmatively, it is not clear whether this implies that (a) he does in fact hold the attitude of hope toward this content, or only that (b) he believes that he holds the attitude of hope toward this content. In responding affirmatively, does the subject hold the attitude of

hope toward the content presented by Σ, or does he merely hold the attitude of belief toward the content that he holds the attitude of hope toward the content expressed by Σ? I am not sure how to answer this question.

13 The "wide" and "narrow" views of rationality are discussed in greater detail Ruth Barcan Marcus 1990.

14 I am not maintaining that if a subject assents to an occurrence of Σ, then he is conscious of his belief that this occurrence is true. In most cases it seems that the assenting subject would not be aware that he has a meta-linguistic belief to the effect that the relevant occurrence is true, but it does not follow from this that the assenting subject has no such meta-linguistic belief. If I eat a sandwich, it follows, other things being equal, that I believe that the sandwich is edible, though it is unlikely that I am aware of this belief while eating the sandwich.

15 Barwise and Perry 1983.

16 It may be helpful to compare these two kinds of psychological theory with the conceptual framework in Stich 1983. In Stich's terminology, both the above described kinds of psychological theories--intentional and non intentional--are *folk theories*; i.e. on either kind of theory the terms 'belief' and 'desire' will have designations, and will not go the way of 'phlogiston'. The distinction between the two kinds of theories concerns whether or not the intentional notions of belief, desire, etc. are *reducible* to non intentional notions. (Note that if Stich's skeptical thesis is correct, if the notions of belief and desire will play no role at all in a comprehensive psychological theory, then neither a Fregean theory nor a Russellian theory will be able to be incorporated into a comprehensive psychological theory.)

17 Fodor "Precis of Modularity of Mind," in Fodor 1990.

18 If we assume, following Higginbotham, that linguistic knowledge, including knowledge of the semantic values of terms, is propositional knowledge which is somehow represented in the brain, then it is not at all surprising that the perception of occurrences of distinct types results in distinct behaviors. The input to the linguistic module is a representation of an occurrence. This representation is processed by the linguistic component, and a representation of a proposition is derived. In deriving this propositional representation, only relevant pieces of lexical knowledge, knowledge of the semantic values of terms, are utilized. Thus the processing of an instance of a mediating state which represents (O1) will involve the utilization of different pieces of lexical knowledge than will the processing of an instance of a mediating state with represents (O2). It also plausible to suppose that which pieces of lexical knowledge get utilized affects which pieces of more general knowledge get utilized--the lexical knowledge governing 'Twain' may be "hooked up" to pieces of knowledge to which the lexical knowledge governing 'Clemens' is not "hooked up." Hence, contingent upon the mediating state input, entirely different aspects and abilities of one's mind may be utilized and it is no surprise that the behavior, the output, is distinct.

Also note that, following a suggestion of Gabrial Segal's, if the above picture is close to accurate, and the business of Semantics is to account for what goes on in part of a speaker's linguistic module, then it may not be the business of Semantics to account for the cognitive significance of occurrences. That which explains the cognitive significance of occurrences may not be a feature of the linguistic module at all.

[19] A philosopher, such as myself, would however be within his bounds if he were to point out that, however σ and σ' are in the end to be individuated, they will be individuated extremely finely, and thus cannot be mapped one-to-one with the *contents* expressed by occurrences, as contents must be public entities.

[20] My views owe a considerable debt to the views expressed in the following: Heim 1982 and Heim 1983; Kamp 1981, Kamp 1985, Kamp 1988 and Kamp 1990; Stalnaker 1979, Stalnaker 1981, Stalnaker 1987a Stalnaker 1987b, and Stalnaker 1988.

[21] Cases of accommodation are discussed by Lewis 1979a.

[22] Though my account of the truth conditions of an occurrence owes a great deal to Kamp's theory, there are several important differences between our views. Two of these differences are pointed out in the footnotes that follow.

[23] There are obvious similarities between the meta-linguistic proposition that 'Twain is a great author' relative to context c presents a truth, and what Stalnaker calls the "diagonal proposition" which is determined by an occurrence of 'Twain is a great author'. The diagonal proposition determined by an occurrence is, roughly, the set of possible worlds w such that if the occurrence took place in w, it would be true in w.

[24] Contexts, as I understand them, do not serve as what Kaplan would call "contexts of evaluation," as contexts do not contain possible worlds.

[25] This example is based upon the famous "Hob Nob sentences" first discussed in Geach 1962.

[26] For a more detailed discussion of these constraints, see Heim 1982, and Heim 1983.

[27] An intuitively adequate definition of truth for an occurrence would at least resemble the following:

> An token τ of sentence Σ uttered relative to DRS D which induces a DRS D' is true iff there is an appropriate assignment function $f( )$ such that (i) $f( )$ satisfies all conditions of D' that are *relevant to* τ, and (ii) $f( )$ obeys all internal and external restrictions of D' that are relevant to τ.

What is it for a condition or restriction of a DRS D to be *relevant to* a token τ? Roughly, a condition or restriction is *relevant to* a token τ just in case the condition or restriction

contains a reference marker which is linked to a reference marker which corresponds with an occurrence of an NP appearing in τ. (Note, again, that if there is no appropriate assignment function, the token τ relative to D is not true; it does not follow that the occurrence is false.) This definition seems to work for discourses containing occurrences of simple declarative sentences, but complications arise when the definition is applied to discourses containing conditionals and other complex sentences. The difficulties here arise because providing an intuitively adequate definition of truth for an occurrence requires that one have a solution to the projection problem, and I have no solution to this general problem.

28 My analysis of occurrences of belief ascriptions differs from Kamp's in that I do not posit reference markers which represent belief states. (Asher presents an analysis very similar to Kamp's in Asher 1986.) Under Asher and Kamp's analysis, condition (4) of DRS-2b would read (something like)

>   (4*) believes <a, p> and p is *characterized** by the following sub-DRS:

>   Universe: c

>   Conditions: (1) c is a great author.

Reference marker p represents a belief state. Thus Kamp and Asher maintain that belief is a relation between agents and belief states. Kamp and Asher seem to differ, however, concerning what it is for a belief state p to be *characterized** by a DRS. Kamp seems to think that a belief state p is *characterized** by a DRS D just in case the set of possible worlds in which p is true is a subset of the worlds in which D is true. For Asher, however, D must also have a "similar structure" to p. I think both Asher and, to lesser extent, Kamp are mistaken in supposing that attitude ascriptions attempt to describe mental states directly. Odile's holding the attitude of belief toward p requires that Odile be in a some kind of belief state, but in reporting that Odile believes such and such, I am not attempting to describe this particular belief state *directly*.

That ordinary attitude ascriptions do not describe, or refer to, mental states directly is one of the lessons learned in Chapter 3.

29 The following constitutes a more or less formal method for determining the propositions determined by a given sub-DRS D under an assignment function $f( )$:

D contains a set S={C_1, C_2, . . . C_n} of conditions. Take the power set of S, P(S). For every set of conditions s element of P(S), follow the following procedure:

>   First--conjoin the conditions in s. (E.g. if the elements of s are 1. b is named 'Santa', and 2. a kissed b, then form the conjunction, 'b is named 'Santa' & a kissed b'.)

Second--replace the reference marker terms with variables (alphabetic variants). (Continuing with our example, from the conjunction obtain, 'x is named 'Santa' & y kissed x'.)

Third--Bind the left-most variable with a lambda operator. (Thus we obtain 'λx[x is named 'Santa' & y kissed x]( )'.)

Fourth--Bind the next left-most variable with a lambda operator, and continue until all variables are bound. (Thus we obtain 'λy[λx[x is named 'Santa' & y kissed x]( )]( )'.)

What we have now is an n-ary predicate (where n is the number of distinct variables appearing in the open sentence obtained in step two.) Note that there is a one-to-one correspondence between open argument places, and types of original reference marker terms; each open argument place is associated with a reference marker term type. (E.g. in our complex predicate 'λy[λx[x is named 'Santa' & y kissed x]( )]( )' the inner-most open argument place is associated with the reference marker term type 'b', and the outer-most open argument place is associated with term type 'a'.) What we need to do now is to either plug the appropriate entities into the appropriate argument places of the property, or relation, designated by the n-ary predicate, or bind the argument place with an existential quantifier. These tasks are accomplished in the final two steps:

Fifth--For every open argument place in the n-ary predicate, if this argument place is associated with a reference marker m such that m is linked to a reference marker in the universe of the primary DRS, then plug a name referring to $f(m)$ into the open argument place. If this argument place is associated with a reference marker m which is not linked to a reference marker in the universe of the primary DRS, then go to step six. (Continuing with our example, suppose that $f(a)$ = Sam, and that reference marker b is not linked to a reference marker in the universe of the primary DRS. Thus in step five 'Sam' is plugged into the "outer-most" open argument place of our predicate, and we obtain 'λy[λx[x is named 'Santa' & y kissed x]( )](Sam)'.)

Sixth--For all argument places which are associated with a reference marker m such that m is not linked to a reference marker in the universe of the containing DRS, each open argument place is to be filled with an appropriate variable, and bound--from the outside--by an existential quantifier. Thus, continuing with our example, the inner-most open argument place in our predicate is bound--from the outside--by an existential quantifier. Thus we obtain:

'$\exists z$ [λy[λx[x is named 'Santa' & y kissed x](z)]](Sam)'

The proposition presented by the resulting closed sentence is the proposition s contributes toward the propositions determined by sub-DRS D under assignment function $f(\ )$.

Finally, the set of propositions determined by sub-DRS D under $f(\ )$ is the set

$$\{p:\ p\text{ is determined by s under }f(\ )\text{ where s is a member of P(S)}\}.$$

30 On my view the *de re/de dicto* distinction is a distinction between two kinds of ascription occurrences, and not a distinction between kinds of propositional attitudes. On my view there is no such thing as a *de dicto*, or *de re*, *belief*. (Though I suppose it might be said that on my view all attitudes are attitudes *de re*, but I think this would be misleading way to put it. Also, on my view 'believes' is not ambiguous between "believes *de dicto*," and "believes *de re*.")

31 I believe that the term '*de se*' is due to Lewis 1979b.

32 For a more formal account of how DRSs are constructed from discourses, see Kamp 1981.

33 Both Kamp and Heim seem to think that DRT will *supplant* traditional truth theoretic semantics. However, there is reason to doubt this conception of DRT: It not at clear that DRT is able to account for the truth conditions of "donkey sentences," as Heim and Kamp originally claimed it could. The difficulties DRT has with "donkey sentences" seem to arise because Heim and Kamp's (rather complicated) analyses of conditionals and quantified sentences are fundamentally incorrect. What I am suggesting is that these shortcomings of DRT may not pose problems for a theory which combines DRT with a more traditional truth-theoretic semantics.

34 For serious attempts to address the first two concerns, see Kamp 1981, and Asher 1986. For a serious attempt to address the third issue, see Kamp 1990. (I should point out that in this latter paper Kamp does not endorse a Russellian analysis of the attitude of belief.)

Appendix A. The Crux of the Matter: Cognitive Values and the Phenomenon of Recognition Failure.

In this appendix I will argue that, regardless of Charybdis, Fregean theories cannot steer clear of Scylla; i.e. Fregean theorists cannot provide a principle of individuation for cognitive values which satisfies both the individuation constraint and the independence constraint. For the sake of illuminating what I think are interesting and important features of the phenomenon of opacity, however, the argument in this appendix will take a rather circuitous route: In section A.1 I will first illustrate that the essence of the phenomenon of opacity is the phenomenon of *recognition failure*, and I will restate the Scylla facing the Fregean theorist in terms of the phenomenon of recognition failure. That is, I will illustrate that if the Fregean theorist is to provide an adequate theory of the phenomenon of opacity, he must provide an adequate theory of recognition failure, and this, in turn, requires that the Fregean theorist provide a principle of individuation for cognitive values which satisfies both the individuation constraint, and the independence constraint, as these constraints pertain to the phenomenon of recognition failure. I will then introduce three kinds of individuation principles for cognitive values, and argue that any principle which is of one of these kinds will fail to satisfy either the individuation constraint or the independence constraint. Finally, in section A.2 I will explicate what I take to be the conceptual foundations of the Fregean strategy and I will argue that the Fregean strategy for explaining the phenomenon of recognition failure is fundamentally misconceived.

A.1 Cognitive Values and the Phenomenon of Recognition failure.

The essence of the arguments from opacity is the phenomenon of recognition failure. The reason that Odile assents to (O1), yet dissents from (O2) is that she in some sense does not *recognize* Clemens as Twain; if there is an instance of opacity involving

Odile, (O1) and (O2), then she must not believe that the tokens of 'Twain' and 'Clemens' appearing in these occurrences are coreferential. Thus I conclude that in general, if, in a context c, a subject assents to an occurrence such as (O1), yet dissents from an occurrence such as (O2), then there is some identity sentence such as

(44) Twain is identical to Clemens.

such that, relative to c, the subject is disposed to dissent from occurrences of this sentence.

Let us say that a normal understanding subject is in a *state of recognition failure* concerning $\ulcorner \alpha$ is identical to $\beta \urcorner$ relative to a context c, just in case he is, relative to c, disposed to dissent from occurrences of $\ulcorner \alpha$ is identical to $\beta \urcorner$, where α and β are coreferential relative to c.[1] Furthermore, let us say that any situation in which a subject is in a state of recognition failure is an *instance of the phenomenon of recognition failure*. The following nomological conditional can now be formulated:

<u>Nomological Conditional 1</u>: If in context c there is an instance of opacity involving a subject x, and terms α and β, then, *ceteris paribus x* is in a *state of recognition failure* concerning $\ulcorner \alpha$ is identical to $\beta \urcorner$ relative to c.

(The conditional is not a material conditional; rather it is nomological. A nomological conditional is, roughly, a material conditional which holds in all possible worlds nomologically similar to the actual world, but does not hold in all possible worlds.)

The converse of Nomological Conditional 1 is also true. I assume that any subject who understands occurrences of referring term α relative to a context c is not in a state of recognition failure concerning occurrences of $\ulcorner \alpha$ is identical to $\alpha \urcorner$ relative to c. Granted this assumption, the following nomological conditional also holds:

<u>Nomological Conditional 2</u>: If there is, in a context c, an *instance of recognition failure* concerning to $\ulcorner\alpha$ is identical to $\beta\urcorner$ involving a subject x, then, *ceteris paribus*, there is a sentential function $\Sigma(\ )$ such that, relative to c, x is disposed to assent to occurrences of $\Sigma(\alpha)$, yet dissent from occurrences of $\Sigma(\beta)$.

If we allow $\Sigma(\ )$ be a sentential function of the form, $\ulcorner\alpha$ is identical to __$\urcorner$ the conditional is almost trivially true: If $\Sigma(\ )$ is of the form $\ulcorner\alpha$ is identical to __$\urcorner$, then $\Sigma(\alpha)$ is of the form $\ulcorner\alpha$ is identical to $\alpha\urcorner$ and $\Sigma(\beta)$ is of the form $\ulcorner\alpha$ is identical to $\beta\urcorner$. In order to satisfy the antecedent x must be disposed, relative to c, to dissent from occurrences of α is identical to β and it was assumed above that any subject who understands α relative to c is disposed to assent to occurrences of $\ulcorner\alpha$ is identical to $\alpha\urcorner$ relative to c. I claim, however, that the conditional holds in nontrivial cases as well; i.e. I claim that the conditional is true for some $\Sigma(\ )$ not of the form $\ulcorner\alpha$ is identical to __$\urcorner$. If, relative to a context c, Odile is disposed to dissent from an occurrences of

(44) Twain is identical to Clemens.

there must be some reason why Odile does not believe that these tokens of 'Twain' and 'Clemens' are coreferential; there must be some property, e.g. *being witty*, such that Odile believes that the referent of the token of 'Twain' has this property, while she does not believe that the referent of the token of 'Clemens' has it. Consequently, if there is, in a context c, an instance of recognition failure involving Odile and an occurrence of (44), then there must be some predicate Π such that, relative to c, Odile is disposed to assent to occurrences of

(45) Twain is Π.

though she is disposed to dissent from occurrences of

(46) Clemens is Π.

Thus the Nomological Conditional 2 is true in nontrivial cases in which Σ( ) is not of the form $\ulcorner \alpha$ is identical to ___$\urcorner$ as well.

I assume that it follows from the fact that a subject is *disposed*, relative to a context c, to perform a certain action that there is a possible world and a possible context c' in which he does perform this action. If this assumption is correct, it then follows from Nomological Conditional 2 that

> Nomological Conditional 3: If there is, in a context c, an *instance of recognition failure* concerning $\ulcorner \alpha$ is identical to $\beta \urcorner$ involving a subject x, then, *ceteris paribus*, there is a nomologically possible instance of opacity involving α, β, and x.

If we now combine Nomological Conditional 1 with Nomological Conditional 3 the following nomological biconditional is derived:

> The Law of Recognition Failure: A subject x is in a *state of recognition failure* concerning $\ulcorner \alpha$ is identical to $\beta \urcorner$ relative to a context c IFF, *ceteris paribus*, relative to c, there is a nomologically possible instance of opacity involving α, β, and x.

The Law of Recognition Failure can be invoked in order to formulate arguments from opacity against the Naive Russellian Theory. For example, let α be replaced by 'Twain' and β be replaced by 'Clemens'; if, relative to some context c, Odile is in a state of recognition failure concerning 'Twain is Clemens', then it follows that there is a nomologically possible instance of opacity in which Odile assents to an occurrence of Σ('Twain'), yet dissents from an occurrence of Σ('Clemens'). Since the Naive Russellian Theory maintains that occurrences of 'Twain' and 'Clemens' express the same content, this nomologically possible instance of opacity is an instance of opacity with regard to

the Naive Russellian Theory. Therefore it follows from the Principle of Arguments from Opacity, i.e.

> The Principle of Arguments from Opacity: For all theories T, there is a nomologically possible instance of opacity with regard to T iff, *ceteris paribus*, corresponding epistemological and semantic arguments from opacity against T can be formulated.

that corresponding epistemological arguments from opacity can be formulated.

The point of going through the rather tedious process of formulating the Law of Recognition Failure is to justify and make precise the claim that the essence of the phenomenon of opacity is the phenomenon of recognition failure. Arguments from opacity against the Naive Russellian Theory can be formulated because there are instances of recognition failure; that is, the Law of Recognition Failure makes it clear that if there is a nomologically possible instance of recognition failure, then semantic and epistemological arguments from opacity against the Naive Russellian Theory can be formulated. It was shown in section 2.2, however, that it is at least in principle possible to formulate arguments from opacity against Fregean theories. In order to reflect this fact, the Law of Recognition Failure must be amended so that it can be invoked in order to formulate arguments from opacity against Fregean theories as well as the Naive Russellian Theory. This can be done by relativizing the phenomenon of recognition failure to particular theories of content: let a subject be in a state of recognition failure concerning $\ulcorner\alpha$ is identical to $\beta\urcorner$ relative to c *with regard to theory T* just in case he is, relative to c, disposed to dissent from occurrences of $\ulcorner\alpha$ is identical to $\beta\urcorner$, where according to T, α and β express the same content relative to c. The Law of Recognition Failure can then be amended so that it applies to Fregean theories as well as Russellian theories as follows:

> The Relativized Law of Recognition Failure: A subject x is in a state of recognition failure concerning $\ulcorner\alpha$ is identical to $\beta\urcorner$ *with regard to* T relative to a context c IFF, *ceteris paribus*, relative to c, there is a nomologically possible instance of opacity with regard to T involving α, β, and x.

This more general statement of the Law of Recognition Failure can be invoked to formulate arguments from opacity against Fregean theories as well as the Naive Russellian Theory: if there is a sentence of the form $\ulcorner\alpha$ is identical to $\beta\urcorner$ such that according to the Fregean theory F, relative to c, α and β express the same cognitive value, and there is a subject who is disposed, relative to c, to dissent from occurrences of $\ulcorner\alpha$ is identical to $\beta\urcorner$, then, by the Relativized Law of Recognition Failure, there is an instance of opacity with regard to F. If we now invoke the Principle of Arguments from Opacity,

> The Principle of Arguments from Opacity: For all theories T, there is a nomologically possible instance of opacity with regard to T iff corresponding epistemological and semantic arguments from opacity against T can be formulated.

then both semantic and epistemological arguments from opacity against Fregean theory F can be formulated.

It follows from the Relativized Law of Recognition Failure that the Fregean strategy for precluding the arguments from opacity can succeed if and only if there is some Fregean theory F which individuates cognitive values so finely that it is nomologically impossible for there to be *an instance of recognition failure* with regard to F. Therefore, Fregean theories can be, and should be, understood as endeavoring to account for the phenomenon of recognition failure; they can be understood as endeavoring to provide an *explanation* as to why, e.g., Odile might be disposed to dissent from occurrences of

(44) Twain is identical to Clemens.[2]

Both the Individuation Constraint and the Independence Constraint can be restated in terms of the phenomenon of recognition failure: An adequate Fregean theory F must satisfy, first,

> The Individuation Constraint: It is nomologically necessary that if there is an instance of recognition failure concerning $\ulcorner\alpha$ is identical to $\beta\urcorner$ involving a subject x, then F distinguishes the cognitive values expressed by these occurrences of α and β for x.

and second,

> The Independence Constraint: The cognitive values posited by F must be individuated wholly independently of the phenomena of recognition failure and opacity; it must be in principle possible to state the individuation conditions for cognitive values without appealing to the phenomenon of recognition failure, or the phenomenon of opacity.

The challenge facing the Fregean theorist is to provide a principle of individuation for cognitive values which satisfies both of the above constraints. I claim that no such principle of individuation has yet been offered; I claim that for every proposed Fregean theory F, either (i) F does not even attempt to provide a principle of individuation for cognitive values,[3] or (ii) F does not individuate cognitive values finely enough--there is a nomoiogically possible instance of recognition failure with regard to F, or (iii) the principle of individuation offered by F itself appeals to instances of the phenomenon of recognition failure, or the phenomenon of opacity.[4] In what follows I will consider three kinds of individuation principles, and I will argue that any principle of one of these kinds will fail to satisfy either the individuation constraint, or the independence constraint, as stated above.

315

I will consider three kinds of individuation principles for cognitive values: *phenomenologically oriented* principles of individuation, *description/property oriented* principles of individuation, and finally principles of individuation which appeal to the notion of *conceptual role*. I will argue that *phenomenologically oriented* and *description/property oriented* principles violate the individuation constraint, while principles which appeal to the notion of *conceptual role* violate the independence constraint.

1. *Phenomenologically oriented* principles of individuation. One might attempt to individuate cognitive values expressed by occurrences of referring terms α and β by appeal to phenomenological properties; the cognitive value expressed by an occurrence of a term for a subject might be identified with kinds of phenomenological experience that the subject associates with the referent of the term. For instance, the Fregean theorist might attempt to explain Odile's dissent from an occurrence of

(44) Twain is identical to Clemens.

by claiming that Odile associates with the occurrence of 'Twain' certain phenomenological properties, particular kinds of phenomenological Twain experiences, and associates with the occurrence of 'Clemens' distinct phenomenological properties, or distinct kinds of phenomenological Twain experiences. (Never mind what it is for a kind of phenomenological experience to be *associated* with an occurrence of a term, and never mind what the identity conditions for *kinds of phenomenological experiences* might be.)

The difficulty with such phenomenologically oriented proposals is that they fail to satisfy the individuation constraint. That is, for any Fregean theory F which invokes such a phenomenological principle of individuation, it is possible that there is an instance of recognition failure concerning an occurrence of $\ulcorner \alpha$ is identical to $\beta \urcorner$ where the subject associates with the tokens of α and β the same kinds of phenomenological experiences.

The problem, in essence, is that phenomenological differences are not essential to the phenomenon of recognition failure.

Suppose that Odile perceives Twain at t1 and in so doing undergoes a phenomenological experience of kind K. At t2 Odile meets Twain again, and again she undergoes a phenomenological experience of kind K; her two experiences are phenomenologically identical. (Twin-Earth may have to be invoked in order to explain how this could occur, but that it is not problematic, as long as it *nomologically possible* that there be a Twin-Earth.) Now at t3 Odile is told that the first person she perceived is named 'Twain' and she is told that the second person she perceived is named 'Clemens'. (She is not told, nor is it in any way suggested, that she perceived distinct people at t1 and t2.) The Fregean theorist who endorses a phenomenological approach to individuating cognitive values must maintain that it is *nomologically necessary* that Odile now be disposed to assent to an occurrence of

(44) Twain is identical to Clemens.

because Odile must associate the same kind of phenomenological experience with both 'Twain' and 'Clemens'. It is clear, however, that it is *not* nomologically necessary that Odile be disposed to assent to an occurrence of (44); despite the phenomenological similarity of her perceptual experiences, Odile may have all sorts of reasons for believing that she perceived different people at t1 and t2. Consequently, principles of individuation which appeal to phenomenological properties and/or experiences violate the individuation constraint.

The Fregean theorist might object that in the above argument, I have not even attempted to state what it is for two phenomenological experiences to be of the same kind, and that I have assumed that in perceiving Twain at t1 and t2 Odile has undergone the same kind of phenomenological experience. The Fregean may claim that there is a

317

way of individuating kinds of phenomenological experiences such that Odile undergoes

distinct kinds of phenomenological experiences. I concede to this objection that the

above argument is really an argument schema; until the Fregean theorist provides some

account of what it is for two phenomenological experiences to be of the same kind, no

substantial conclusion can be drawn from the above described example. It is, however,

incumbent upon the Fregean theorist to provide such a principle of individuation; until

the Fregean theorist provides an independent *criterion of identity* for cognitive values, the

Fregean theory has no explanatory power whatsoever. What the above argument, or

argument schema, illustrates is that if a Fregean theory F *does* provide a criterion for the

identity of phenomenological experience kinds, it will be possible to construct a

nomologically possible instance of recognition failure with regard to theory F.[5]

I conclude therefore that no phenomenologically oriented principle of

individuation for cognitive values can succeed in satisfying the individuation constraint.

The essence of the difficulty is that is sufficient for an understanding subject to dissent

from an occurrence of $\ulcorner \alpha$ is identical to $\beta \urcorner$ that the subject not believe that the

occurrences of α and β refer to the same entity, yet it is not necessary for the subject's

having lacking this belief that the subject associate distinct phenomenological properties

with the occurrences of α and β. (Similarly, a subject might believe that *sensory*

impressions α and β are of impressions of distinct entities, even though α and β are as

phenomenologically similar as one likes.)

2. *Description oriented* principles of individuation. Description oriented

principles of individuation maintain that the cognitive value expressed by an occurrence

of a referring term is a cluster of properties, or more precisely a complex property, which

is instantiated by the referent of the term. Thus the Fregean theorist who proposes a

description oriented principle of individuation for cognitive values maintains that Odile

dissents from an occurrence of

(44) Twain is identical to Clemens.

because she associates property P with the occurrence of 'Twain' and associates distinct property P' with the occurrence of 'Clemens'. If this kind of individuation principle is to have any hope at all of satisfying the individuation constrain, P and P' cannot be ordinary general properties which can obviously be believed to be instantiated by more than one thing. For instance, P and P' cannot be identified with the general properties, *is a witty bearded man who smokes a pipe and is a great author*, and *is not a witty bearded man who smokes a pipe and is a great author*, respectively. For it is clear that a subject could associate the same general property with the relevant occurrences of 'Twain' and 'Clemens', and nonetheless believe that these occurrences are not coreferrential; Odile might associate the general property, *is a witty bearded man who smokes a pipe and is a great author* with both 'Twain' and 'Clemens', and yet believe that the referent of 'Twain' is not identical to the referent of 'Clemens'--Odile might think that there are two witty bearded men who smoke pipes are great authors. If Odile were in this state, she would be disposed to dissent from occurrences of

(44) Twain is identical to Clemens.

thus violating the individuation constraint.

The above considerations suggest that if a description oriented principle of individuation for cognitive values is to succeed, the properties appealed to cannot be ordinary general properties, but rather must be *individual properties*: every ordinary general property P determines an individual property P_1 where P_1 is true of an object x just in case x the *unique* entity which satisfies P. If the properties subjects associate with occurrences of terms are identified with such *individual properties*, then it is at least plausible to suppose that a description oriented principle of individuation could satisfy

the individuation constraint: *if* it is granted that no subject can ever believe that there are *two* entities which have the *same* individual property, then whenever a subject associates the same individual property with terms α and β, the subject will believe that α and β are coreferential. The problem is that there is no reason for supposing that a subject could not believe that *two* entities have the *same* individual property; the subject may associate P_1 with both α and β, and yet not realize that he associates the same individual property, P_1, with both α and β. For example, Odile might associate P_1 with both 'Twain' and 'Clemens', and yet not believe that these terms are coreferential. A plausible candidate to serve as the individual property Odile associates with 'Twain' might be *the witty bearded American author who smokes a pipe and lived in Hannibal.* Suppose Odile associates this one individual property with both 'Twain' and 'Clemens'. Why might Odile nonetheless not believe that 'Twain' and 'Clemens' are coreferential? The problem is that Odile might think that there are *two* Hannibals, both of which is home to a witty bearded American author who smokes a pipe. In this case Odile would, despite the fact that she associates the same individual properties with both 'Twain' and 'Clemens', not believe that 'Twain' and 'Clemens' are coreferential. As a result she would dissent from occurrences of

(44) Twain is identical to Clemens.

and thus the independence constraint is violated.

The defender of description oriented principles of individuation might respond to this objection by simply maintaining that subjects cannot *unknowingly* associate the same individual property with distinct referring terms α and β.[6] In terms of the above example, the Fregean theorist may claim that if Odile thinks there are *two* Hannibals, then she does not associate the same individual properties with 'Twain' and 'Clemens'. Odile, the Fregean claims, associates with 'Twain' the individual property *the witty*

American author who lives in Hannibal$_1$, and associates with 'Clemens' the distinct

individual property, *the witty American author who lives in Hannibal$_2$*. And these

individual properties are distinct because Odile does not believe that Hannibal$_1$ is

identical to Hannibal$_2$.

There are two problems with this response. First, and most importantly, in giving

this response the Fregean theorist violates the independence constraint: The Fregean

theorist must provide a principle of individuation for cognitive values which does not

appeal to the phenomenon of recognition failure itself. The Fregean theorist has

identified cognitive values with *individual properties*, and he now proposes that α and β

express the same individual property P_1 for a subject x, only if x is not in a state of

recognition failure concerning P_1, or some constituent of P_1. More precisely, the Fregean

theorist maintains that the cognitive values, the individual properties, expressed by α and

β are the contents expressed by definite descriptions *the* Π and *the* Π'. Moreover, the

contents expressed by these definite descriptions are identical only if there are no

constituent terms τ and τ' of predicates Π and Π' such that x is in a *state of recognition

failure* concerning occurrences of $\ulcorner\tau$ is identical to $\tau'\urcorner$. Thus in giving this response the

Fregean theorist may succeed in individuating cognitive values in accordance with the

individuation constraint, but in accomplishing this much he violates the independence

constraint.

The second problem with the Fregean response is that it seems to engender an

infinite regress: What, exactly, are *Hannibal$_1$* and *Hannibal$_2$*? What are the constituents

which serve to differentiate the individual properties Odile associates with 'Twain' and

'Clemens'? They cannot be cities; there is only one relevant city, Hannibal. Thus, it

seems, the property constituents expressed by '*Hannibal$_1$*' and '*Hannibal$_2$*' must be *ways

of thinking* of Hannibal; and therefore they must be distinct individual properties which

are instantiated by the city Hannibal. But once this move is made, the regress is

imminent: If cognitive values are identified with individual properties, yet individual

properties contain other individual properties as constituents, then the Fregean theorist will be unable to provide a principle of individuation for cognitive values.

3. Principles of individuation which appeal to the notion of *conceptual role*.[7]
Fregean theories which invoke the notion of *conceptual role* presuppose a *representationalist* picture of the mind. For the sake of convenience, I will assume that he relevant mental representations are expressions in *the language of thought*, and I will use capital letters to designate such mental expressions (types). *Conceptual roles* are properties of expressions of mentalese (types). Conceptual role theorists propose that Odile dissents from an occurrences of

(44) Twain is identical to Clemens.

because TWAIN and CLEMENS, the "mentalese correlates" of 'Twain' and 'Clemens', have distinct *conceptual roles*. What is it for two expression (types) of mentalese to have distinct, or identical, conceptual roles?

Though Block does not propose a precise account of the identity conditions of conceptual roles[8], he does describe what such an account might be like:

> conceptual role . . . is a matter the causal role of the expression in reasoning and deliberation and, in general, in the way the expression combines with and interacts with other expressions so as to mediate between sensory inputs and behavioral outputs. A crucial component of a sentence's conceptual role is a matter of how it participates in inductive and deductive inferences. A word's concptual role is a matter of its contribution to the role of sentences.

For example, consider what would be involved for a symbol in the internal representational system, '→', to express the material conditional. The '→' in 'FELIX IS A CAT → FELIX IS AN ANIMAL' expresses the material conditional if, for example, when the just quoted sentence interacts causally with:

'FELIX IS A CAT', the result is a tendency to inscribe 'FELIX IS AN ANIMAL' (other things equal, of course).
'FELIX IS NOT AN ANIMAL', the result is a tendency to prevent the inscription of 'FELIX IS A CAT', and a tendency to inscribe 'FELIX IS NOT A CAT'.
'IS FELIX AN ANIMAL?', the result is a tendency to initiate a search for 'FELIX IS A CAT'.[9]

How, then, are the *conceptual roles* of TWAIN and CLEMENS, for Odile, to be distinguished? These two expressions of mentalese have distinct conceptual roles just in case instantiations of these expressions in Odile's brain have distinct "likely inferential antecedents and consequents," and/or distinct likely "behavioral effects (or at least impulses in motor-output neurons)."[10] Suppose, however, that Odile has very few beliefs concerning the referent of 'Twain' and the referent of 'Clemens'. Suppose that Odile does not believe that Twain is Clemens; she dissents from occurrences of 'Twain is identical to Clemens. Further suppose, however, that for every property P--aside from properties such as, *is named 'Twain'*--Odile believes that both Twain and Clemens have P; Odile does not believe that Twain is identical to Clemens, yet she believes that Twain is a great author, and that Clemens is a great author; that Twain is dead, and that Clemens is dead; that Twain is American, and that Clemens is American, etc. Thus, for (almost) every *mentalese* sentential function $\Sigma(\ )$, $\Sigma(\text{TWAIN})$ and $\Sigma(\text{CLEMENS})$ have the same likely inferential and behavioral antecedents and consequents. For example if $\Sigma(\ )$ is,

____ IS A WITTY AUTHOR, then "inscriptions" of $\Sigma(\text{TWAIN})$ and $\Sigma(\text{CLEMENS})$ are equally likely to have "inscriptions" of THERE IS AT LEAST ONE WITTY AMERICAN AUTHOR as consequents.

Given Odile's beliefs, how can the conceptual role theorist distinguish the conceptual roles of TWAIN and CLEMENS for Odile? The conceptual role theorist might attempt to distinguish the conceptual roles of TWAIN and CLEMENS for Odile by appeal to the sentential function, _____ IS IDENTICAL TO TWAIN. Given the above description of Odile's beliefs, an "inscription" of TWAIN IS IDENTICAL TO TWAIN, will have a very different conceptual role than CLEMENS IS IDENTICAL TO TWAIN: an "inscription" of the former, but not the latter is likely play a direct role in causing Odile to assent to an occurrence of

(44) Twain is identical to Clemens.

The difficulty of course is that in individuating the conceptual roles of TWAIN and CLEMENS in this way, the conceptual role theorist violates the independence constraint: if the cognitive role theorist is to explain why Odile dissents from an occurrence of (44) by appeal to (alleged) fact that TWAIN and CLEMENS play distinct conceptual roles, then he cannot individuate the conceptual roles of these mentalese expressions by appeal to the fact that Odile dissents from, or is disposed to dissent from, an occurrence of (44).[11]

Block is well aware of the threat of circularity. As Block himself states,

> The [conceptual role theorist's] explanation of behavior may seem circular, hence trivial. How can [he] characterize a [conceptual role] . . . in terms of a tendency for representations that have it to cause jumping, [or dissenting] and then turn around and explain jumping [or dissenting] by appeal to a representation's having this [conceptual role]?[12]

Block's principle response to this difficulty is to point out that "a functionally individuated entity can, in principle, be identified by independent (usually physicalistic) means and the mechanism of its causal connection to the effects described. For example,

324

a gene identified functionally via the methods of Mendelian genetics can be identified as a clump of DNA via the methods of molecular genetics."[13] I, of course, concur with Block's point, viz. there would be no circularity in the conceptual role theorist's explanation of behavior if conceptual roles, following the paradigm of Mendelian genes, could be individuated by appeal to independent physicalistic properties. But I do not understand how conceptual roles thus individuated would properly called "conceptual roles." That is, if behavior could be explained in physicalistic terms by appeal to physical states and/or properties, without appeal to behavioral effects, then what explanatory work could the notion of *conceptual role* do? There are not kinds of DNA *in addition to* genes; rather genes *are* kinds of DNA. To vindicate the explanatory relevance of conceptual roles by providing purely physicalistic individuation conditions for them would be to render the notion of *conceptual role* superfluous.[14]

Perhaps then, cognitive values should be identified with physicalistically individuated states and/or properties; perhaps a physicalistic principle of individuation which satisfied both the individuation constraint and the independence constraint could be formulated. If the sole task were to explain behavior, e.g. explain why Odile dissents from a particular occurrence of (44), then this empirical strategy would, I believe, be the most promising avenue to pursue. But it is not at all clear that this avenue is open to the Fregean theorist. The Fregean theorist maintains that the phenomena of recognition failure and opacity can be explained by appeal to *meanings*, but surely publicly accessible *meanings* cannot be identified with complex physicalistically individuated properties or states. If the Fregean theorist concedes that the phenomena of recognition failure and opacity can be explained only in terms of complex physicalistically individuated properties or states, then he must concede that these phenomena cannot be explained by appeal to *meanings*. But if the Fregean theorist concedes this much, then he concedes that there is no reason to posit cognitive values over and above referents.

A.2 The Misconceived Fregean Model of Recognition Failure.

I maintain that Fregean theories are unable to provide adequate principles of individuation for cognitive values because the Fregean strategy for explaining the phenomenon of recognition failure is fundamentally misconceived. Consider the following famous passage from "Uber Sinn und Bedeutung":

> The sense of a proper name is grasped by everybody who is sufficiently familiar with the language or totality of designations to which it belongs; but this serves to illuminate only *a single aspect of the referen[t]*, supposing it to have one. Comprehensive knowledge of the referen[t] would require us to be able to say immediately whether any given sense belongs to it. To such knowledge we never attain, (my emphasis).[15]

Frege maintains that every "ordinary referent" has associated with it a (probably infinite) number of "aspects" or senses. To think about a referent, to entertain a thought about a referent, it is necessary to *grasp* one of its aspects. Because associated with any referent there are many aspects, it is possible for a subject to grasp two distinct aspects of the same referent. It is this possibility which is supposed to account for recognition failure: According to Frege's theory of sense and reference, recognition failure can occur only if, for example, one grasps the aspect associated with 'Twain', and also grasps the aspect associated with 'Clemens', and these aspects are not identical, though they are both aspects of the same referent, i.e. both aspects correspond with the man Samuel Clemens. *Recognition failure* concerning (44) is then alleged to be a matter of the subject's *failure to realize* that both of these grasped aspects correspond with Samuel Clemens.

This general model of recognition failure is employed by all Fregean theories: associated with any object there are an infinite number of aspects, or "ways of thinking" of the object. (The aspects, or ways of thinking, of an object of course are the cognitive values presented by terms which refer to the object.) A subject can *fail to recognize* an

object only if the subject "grasps" or in some way cognizes, two aspects of the object and *fails to realize* that both these aspects correspond with the same object. Thus, according to the Fregean model of recognition failure, a subject is in a state of recognition failure concerning ⌜α is identical to β⌝ relative to c, (where α and β are coreferential relative to c), just in case α and β are associated with distinct aspects, or "ways of thinking," of their referent, and the subject fails to realize that these aspects correspond with the same referent. Furthermore, it is this *realization failure* which accounts for the subject's disposition to dissent from occurrences of ⌜α is identical to β⌝ relative to c; the subject is disposed to dissent from ⌜α is identical to β⌝ because he *fails to realize* that the cognitive values presented by α and β, relative to c, have the same referent. Thus the essence of the Fregean model of recognition failure, and thus the essence of the Fregean strategy for precluding the arguments from opacity, is to reduce a subject's *failure to recognize* an object to the subject's *failure to realize* that two ways of thinking of that object are ways of thinking of the same object.

In positing aspects, or cognitive values, and reducing the phenomenon of *recognition failure* to the phenomenon of *realization failure*, has the Fregean theorist made explanatory progress? The Fregean theorist is attempting to describe the conditions under which a subject will fail to recognize the referent of α as the referent of β; i.e. he is attempting to explain why a subject might not believe that the referent of α is the referent of β. The Fregean suggests that this will occur whenever (i) the subject associates distinct aspects of the referent with occurrences of α and β *and* (ii) the subject *fails to realize* that these aspects correspond with the same referent. (Note that the second conjunct is crucial: the Fregean theorist must allow that a subject might associate distinct aspects with α and β, and yet realize that they are aspects of the same entity. That is, the Fregean must allow for a subject to believe what is expressed by an *informative* and true statement of identity.) But in reducing the phenomenon of recognition failure to the phenomenon of realization failure, has the Fregean made any

progress at all? We began with the question, "Under what conditions will a subject not believe that the referent of α is the referent of β?" The Fregean theorist has posited cognitive values, and replaced that original question with the question, "Under what conditions will a subject *fail to realize* that the aspect associated with α is an aspect of the same entity as the aspect associated with β?" But why is that subject's sometimes realize that two aspects correspond with the same referent, and other times do not come to this realization? Though no Fregean theorist I know of has even addressed this question, it seems that it is necessary and sufficient for a subject's *failing to realize* that two aspects are aspects of the same entity that the subject *not believe* that the aspects are aspects of the same entity. If this assumption is correct, then the Fregean has managed to reduce the question, "Under what conditions will a subject not believe that the referent of α is the referent of β?" to the question, "Under what conditions will a subject not believe that the aspect expressed by α is an aspect of the same entity as the aspect expressed by β?"

Note, however, that the problems posed by the above two questions are isomorphic: the original problem was to explain how a subject might not believe that α and β correspond, via the reference relation, to the same entity. The new problem is to explain how a subject might not believe that aspect-1 and aspect-2 correspond, via the "is an aspect of" relation, to the same entity. Thus, since the new problem and the old problem are structurally isomorphic, the Fregean theorist must endorse the same strategy in attempting to solve the new problem. That is, he must posit entities, *meta-aspects*, and maintain that a subject can fail to believe that aspect-1 and aspect-2 (which are in fact aspects of the same entity) are aspects of the same entity, if and only if there are distinct *meta-aspects* associated with both aspect-1 and aspect-2, and (i) the subject associates distinct *meta-aspects* of the referent with aspect-1 and aspect-2 and (ii) the subject *fails to realize* that these meta-aspects are meta-aspects of the same referent. But now of course the regress is imminent: If the correct response to the first problem, explaining

recognition failure, is to posit aspects and reduce recognition failure to *realization failure*, then the correct response to the second problem, explaining *realization failure*, must be to posit *meta-aspects*, and reduce the problem of realization failure to the problem of *meta-realization failure*, and so on.

The Fregean theorist may of course object that an infinite regress is not engendered because there is a point at which the Fregean strategy of positing aspects (or meta-aspects, or meta-meta-aspects, etc.) and reducing realization failure (or meta-realization failure, etc.) to meta-realization failure (or meta-meta-realization failure, etc.) is not the correct response. But if the Fregean objects in this manner, then the onus is on him to explain why the Fregean strategy is appropriate with regard to the *first* problem, but not the 1+*x*th problem. If the regress can be stopped, if the Fregean strategy is inappropriate at some level, then why is it appropriate at the first level, as a response to the problem of recognition failure?

It is not surprising, upon reflection, that the Fregean strategy makes no explanatory progress: It seems likely that the most promising procedure for providing an explanation of the phenomenon of *recognition failure* would be to first provide a theory of *recognition success*, and then explain the phenomenon of recognition failure in terms of this theory. Recognition failure would be said to occur whenever one of the necessary and jointly sufficient conditions for *recognition success* did not obtain. The Fregean strategy, however, is utterly vacuous with regard to the phenomenon of recognition success: the Fregean theorist maintains that a subject succeeds in recognizing the referent of α as the referent of β just in case (i) α and β express the distinct aspects, *or one and the same aspect*, and (ii) the subject *realizes* that the aspect associated with α is an aspect of the same entity as the aspect associated with β. In the case recognition success, the posited aspects do no explanatory work whatsoever; recognition success can occur when the relevant terms α and β express the same aspect, or distinct aspects. Why not simplify matters and simply maintain that a subject succeeds in recognizing the

referent of α as the referent of β just in case he realizes that α and β are coreferential? Granted, this is an entirely vacuous explanation, but parsimony dictates that it is to be preferred over the explanation proffered by the Fregean strategy.

The above remarks indicate that the Fregean strategy is fundamentally misconceived. *If* there is a general explanation of the phenomena of recognition failure and recognition success, then it is not the explanation proffered by the Fregean theorist. But why should it even be supposed that there is some *general* explanation of the phenomena of recognition failure and recognition success which can be discovered *a priori*? The phenomenon of recognition failure is the phenomenon whereby a normal, suitably informed, subject does not believe that α and β refer to, or more generally *correspond to,* one thing, where in fact α and β do refer to, or correspond to, one thing. And the phenomenon of recognition success is the phenomenon whereby a normal, suitably informed, subject believes that α and β refer to, or correspond to, one thing, where in fact α and β do refer to, or correspond to, one thing. (In the situations I have been concerned with, 'α' and 'β' have been interpreted as metalinguistic variables ranging over referring terms. But the above remarks apply, *mutatis mutandis*, if 'α' and 'β' are interpreted as variables ranging over perceptual impressions of entities.) Why should it be supposed that there is some *general, a priori,* explanation as to why a subject will, or will not, believe that α and β correspond with the same entity? It is clear that a subject may believe, or not believe, that α and β refer to (or correspond to) the same entity for any number reasons. Why should Frege and his followers have even thought it possible to state an *a priori* and *general* theory of recognition and recognition failure?

The answer, I suggest, is that Frege and his followers have wrongly assumed that *recognition failure* can be explained by appeal to *meanings*, where meanings are publicly accessible, and in some sense knowable *a priori*. If recognition failure could be explained in terms of *meanings*, if a subject's failure to believe that α and β refer to the same entity could be explained in terms of the *meanings* of α and β, then it would be

possible to state a general, *a priori*, theory of recognition and recognition failure. Therefore, if, as I have been arguing, it is *not* possible to state a general, *a priori*, theory of recognition success and recognition failure, then explanations of these phenomena cannot be given by appeal to *meanings*. The Fregean strategy for explaining the phenomena of recognition failure and recognition success is fundamentally misconceived, and the misconception can be traced, I suggest, to the mistaken assumption that these phenomena can be explained by appeal to the *meanings* of public language expressions.[16]

Appendix A Notes

1 For the sake of clarity I define, or characterize, the phenomenon of recognition failure in terms of a subject's disposition to dissent from an occurrence of a sentence of the form $\lceil\alpha$ is identical to $\beta\rceil$. I maintain, however, that the conclusions reached in the section would also apply if the phenomenon of recognition failure were defined so as to included situations in which, where α and β are "coreferential" sensory impressions of some kind, a subject is disposed to behave as if α and β are not "coreferential."

2 Note that if this claim is true, then there is only a superficial difference between the puzzles posed by the phenomenon of opacity, and what others call "Frege's puzzle of identity statements."

3 Most Fregean theorists do no even attempt to provide individuation conditions for cognitive values. Most Fregean theorists are content merely to introduce a term which allegedly refers to cognitive values--"mode of presentation," "mode of apprehension," "guise," appearance," "sense," "representation," "intension," etc.--and to state a *criterion of difference* for cognitive values. Fregean theorists rarely, however, attempt to provide a *criterion of identity* for cognitive values. But without such a criterion of identity, Fregean theories have not explanatory power.

4 Dummett 1978, and following him Evans 1982, have proposed that cognitive values be identified with *abilities to recognize* referents. To the extent that I understand this proposal, I think it is obviously in violation of the independence constraint: cognitive values are what is posited to *explain* one's ability, or one's inability, to *recognize* a referent. Thus, to *identify* cognitive values with abilities is to put the cart before the horse.

5 There are obvious similarities between the argument given here, and the argument against Salmon's theory proffered in section 3.1.2.

6 Here the latent Cartesianism of the Fregean approach manifests itself. If cognitive values are to play roles they were posited to play, then they must be in some way *epistemologically transparent*: cognitive values must be like Cartesian *ideas* in that one cannot "grasp" the same cognitive value twice and *fail to realize* that one has "grasped" the same cognitive values twice. In other words, the Fregean theorist accounts for the phenomenon of recognition failure by positing entities which are such that one *cannot* fail to recognize them, and then analyzing *recognition failure* as a failure to realize that these epistemologically transparent entities correspond to the same referent.

7 In this section I will focus on the "two factor" theory of conceptual role semantics, as sketched in Block 1986, though I think my criticism applies to conceptual role theories in general. I should also point out that though Block does not seem to see his theory as

being *Fregean*, , there are no important differences between Block's "two factor" theory of meaning, and Frege's theory of sense and reference.

8 To his credit, Block states that "*the* crucial question . . . is what counts as identity and difference of conceptual role," Block 1986, p. 628.

9 Block 1986, p. 628. (I do not understand how the single quotes are being used in this passage.)

10 Block 1986, p. 668.

11 The conceptual role theorist may respond that he can distinguish the conceptual roles of TWAIN and CLEMENS, even if Odile's beliefs are as described above. He may claim that if Odile dissents from an occurrence of (44), then there is a mentalese sentential function Σ( ) such that Σ(TWAIN) and Σ(CLEMENS) have distinct conceptual roles, viz. ____ IS NAMED **CLEMENS**. (Words in boldface capitals refer to the mentalese correlates of English quote names; thus CLEMENS is the mentalese correlate of 'Clemens', while **CLEMENS** is the mentalese correlate of "Clemens".) Inscriptions of the sentence of mentalese CLEMENS IS NAMED **CLEMENS** will, it may be assumed, play a role in causing Odile's assent to an occurrence of

(47) Clemens is named 'Clemens'.

while the sentence of mentalese TWAIN IS NAMED **CLEMENS**, it may be assumed, play no such causal role. Conversely, the mentalese sentence TWAIN IS NAMED **TWAIN**, it may be assumed, plays a direct role in causing Odile's assent to

(48) Twain is named 'Twain'.

while the sentence of mentalese CLEMENS IS NAMED **TWAIN**, it may be assumed, plays no such causal role. Therefore it seems that the conceptual role theorist may be able to distinguish the conceptual roles of TWAIN and CLEMENS for Odile without appeal to the phenomenon of recognition failure.

The problem is that the above strategy for individuating the conceptual roles violates the independence constraint because it appeals to the phenomenon of opacity, where the relevant sentential function is either '____ is named 'Clemens", or '____ is named 'Twain". The essence of the phenomenon of opacity is the phenomenon of recognition failure. Thus if the Fregean theorist posits entities or properties to explain the phenomenon of opacity, he can appeal to neither the phenomenon of opacity, nor the phenomenon of recognition failure in individuating the posited entities or properties.

12 Block 1986, p. 668.

13 Block 1986, p. 668.

14 Block seems to be aware of this problem also, but I do not understand his response and I suspect that he is confused. Block responds that,

> The problem has to do with the type-token relation for mental representations. The hope is that there will be a stable physical realization (at least over short stretches of time) of, say, the representation 'CAT', *which of course will be identifiable only by its functional role*. Then, in principle, one could trace the causal links between this representation and behavior, just as the biochemist can in principle trace the mechanism by which a gene affects the phenotype, (my emphasis, Block 1986, p. 669).

What does Block mean when he writes that "the hope is that there will be some stable representation [of CAT] *which . . . will be identifiable only by its functional role.*" That is, how is this statement consistent with Block's previous statement that "a functionally individuated entity *can, in principle, be identified by independent (usually physicalistic) means.*"?

The issue is whether a realization of CAT can be identified *as a realization of CAT* via independent means; the circularity problem is not solved by our being able to point at something in the head of a subject who happens to be thinking of cats. That is, the conceptual role theorist has not solved the circularity problem if he knows that a particular clump of neurons, which he can point to and thereby *identify*, is a realization of CAT only because of its causal consequences. But if we *can* identify the neuron clump *as a realization of CAT* and do so via independent physicalistic means, then "the hope . . . that there will be some stable representation [of CAT] *which . . . will be identifiable* [as a realization of CAT] *only by its functional role*" could not possibly be realized,

Moreover, the gene\DNA analogy hardly serves to support the position Block seems to be defending: scientists *can* identify DNA by purely physicalistic, non-functional, means, and that is why genetics now gives explanations in terms of physicalistically individuated DNA, and not in terms of functionally individuated *genes*.

15 Frege 1956, p. 57-8.

16 Of course it is rather odd to speak of referring terms as sharing *meanings*, or being *synonymous*, in the first place. Consider the adjectives 'remarkable' and the German 'merkwürdig'. These non referring terms are, roughly, *synonymous*. Now consider the referring terms 'Twain' and 'Clemens'. These terms are *coreferential*, but it is difficult to see in what sense they may be said to have the same *meaning*, or be *synonymous*.

Appendix B. Applying DRT to Perry's "Problem of the Essential Indexical," and Richard's "Phone booth Argument."

In this appendix I will illustrate, again rather informally, how DRT can be employed to solve Perry's "problem of the essential indexical," and also to debunk Richard's "phone booth argument."[1]

Perry's so-called "problem of the essential indexical" is really several problems rolled into one; it is partly an epistemological or psychological problem concerning how a subject's behavior is best explained, and it is partly a semantic problem concerning the truth conditions of occurrences. I will here be concerned with the semantic problem only. The semantic problem posed by Perry is a variant of a semantic argument from opacity. Consider the following imaginary situation: We are with Perry at a grocery store, and from behind a two-way mirror we are watching Oscar who is shopping. Oscar notices a trail of sugar on the floor, and we surmise from Oscar's behavior that he believes that somebody has a leaky bag of sugar in their trolley. (Suppose we overhear Oscar saying to the store attendant. 'Sir, somebody has a leaky bag of sugar in their trolley.') At this time Perry says to us,

(49) Oscar believes that a person with a leaky bag is making a mess.

We would judge Perry's utterance of (49) to be true.

Suppose now that Oscar sees in a mirror someone who is pushing a trolley containing a leaky bag of sugar. The person whom Oscar sees in the mirror is actually Oscar, himself, but he does not realize this; Oscar thinks that the person in the mirror with the leaky bag is somebody else. Again, it is apparent to we who are watching Oscar that Oscar does not recognize himself. At this time if Perry were to utter to us

(50) Oscar believes that *he* is making a mess.

(where the occurrence of '*he*' is accompanied by a gesture toward the image in the mirror) our intuitions would again dictate that this utterance is true. Contrary to what is predicted by the Naive Russellian Theory, however, if Perry had uttered

(51) Oscar believes that *he* [himself] is making a mess.

instead of (50), our intuitions dictate that his utterance would have been false. (In order to ensure that occurrences of (50) and (51) present the same Russellian proposition, suppose that the 'himself' appearing in brackets is not voiced by Perry--'himself' as it appears in (51) is merely meant to signify the appropriate interpretation to the reader. Perry's supposed utterance of (51) would thus be an instance of what is sometimes called a "*de se*" belief ascription.") How can DRT account for our intuition that, in the above described situation, an utterance of (50) would be true, while an utterance of (51) would be false?

Suppose that the initial DRS relative to which Perry utters (49), (50) and (51) is as follows:

<u>DRS-8a</u>

Universe: a,

Conditions: 1. a has a leaky bag
2. a is making a mess
3. a is in a belief state which is
characterized by the following sub-DRS:

Universe: b

Conditions: 1. b does not have a leaky bag
2. b is not making a mess

Internal Restriction: $a \leftrightarrow b$

External Restrictions: 1. $R(a)$ = Oscar

DRS-8a depicts the relevant correspondences between Perry's belief state and our belief state(s); we mutually believe that Oscar has a leaky bag and is making a mess, and that Oscar does not believe of himself that he is making a mess.

If Perry now utters an occurrence of

(49) Oscar believes that a person with a leaky bag is making a mess.

relative to DRS-8a, the following DRS would be induced:

<u>DRS-8b</u>

Universe: *a*

Conditions: 1. *a* has a leaky bag
2. *a* is making a mess
3. *a* is in a belief state which is
characterized by the following sub-DRS:

Universe: *b, c*

Conditions: 1. *b* does not have a leaky bag
2. *b* is not making a mess
3. *c* has a leaky bag
4. *c* is making a mess

Internal Restrictions: 1. $a \leftrightarrow b$

External Restrictions: R(*a*) = Oscar

The reference marker *c*, which is introduced by the indefinite NP 'a person with a leaky bag', is added to the universe of the sub-DRS. As it is an indefinite NP, reference marker *c* is not linked to any of the already present reference markers of DRS-8a. The sub-DRS of DRS-8b determines the following Russellian propositions: that Oscar does not have a leaky bag; that Oscar is not making a mess; that somebody has a leaky bag; that somebody is making a mess; and that somebody who has a leaky bag is making a mess. (This last proposition is expressed by '$\exists y (\lambda x[x$ has a leaky bag & *x* is making a mess]$(y))$'. Also, this is not an exhaustive list of the Russellian propositions determined by the sub-DRS of DRS-8b, but it will suffice for our purposes.)

Perry's utterance of

(50) Oscar believes that *he* is making a mess.

338

(where the token of '*he*' is accompanied by a demonstration toward the image of Oscar in the mirror) relative to DRS-8b induces

DRS-8c

Universe: a

Conditions: 1. a has a leaky bag
2. a is making a mess
3. a is in a belief state which is characterized by the following sub-DRS:

Universe: b, c

Conditions: 1. b does not have a leaky bag
2. b is not making a mess
3. c has a leaky bag
4. c is making a mess
5. c's image appears in the mirror.

Internal Restrictions: 1. $a \leftrightarrow b$
2. $a \leftrightarrow c$

External Restrictions: $R(a)$ = Oscar

Because 'he' is a definite NP, the occurrence of 'he' must correspond with a reference marker which is linked to a reference marker already present in DRS-8b. It actually corresponds with a reference marker, viz. c, which is linked to two already present reference markers, viz. c and a. That the occurrence of 'he' corresponds with reference marker c indicates that we take Oscar to believe that person whose image appears in the mirror is the person who has a leaky bag and is making a mess. The internal restriction linking b with a is ensures that the occurrence of 'he' carries a referential presupposition.

Applying our truth definition, the occurrence of (50) relative to DRS-8c is true if and only if there is an assignment function $f(\ )$ such that $f(a) = f(b) = f(c)$ = Oscar; $f(a)$

339

has a leaky bag; $f(a)$ is making a mess; and the following propositions are in Oscar's belief set: $f(a)$ does not have a leaky bag; $f(a)$ is not making a mess; $f(a)$ has a leaky bag; $f(a)$ is making mess; $f(a)$'s image appears in the mirror, and $\lambda x[x$ has a leaky bag & x is making a mess & x's image appears in the mirror](Oscar). (Again, this list of determined propositions is not exhaustive, but it will suffice for our purposes.)

The DRS that would be induced by Perry's utterance of

(51) Oscar believes that *he* [himself] is making a mess.

relative to DRS-8b is, however, slightly different from DRS-8c:

DRS-8c'

Universe: a

Conditions: 1. a has a leaky bag
2. a is making a mess
3. a is in a belief state which is
characterized by the following sub-DRS:

Universe: b, c

Conditions: 1. b does not have a leaky bag
2. b is not making a mess
3. c has a leaky bag
4. c is making a mess
5. b is making a mess.

Internal Restrictions: 1. $a \leftrightarrow b$

External Restrictions: $R(a) = $ Oscar

Applying our truth definition to DRS-8c', Perry's utterance of (51) relative to DRS-8b is true if and only if there is an assignment function $f(\ )$ such that $f(a) = f(b) = $ Oscar; $f(a)$

has a leaky bag; $f(a)$ is making a mess; and the following propositions are in Oscar's

belief state: $f(a)$ does not have a leaky bag; $f(a)$ is not making a mess; $f(a)$ has a leaky

bag; $f(a)$ is making mess; and $\lambda x[x$ does not have a leaky bag & x is not making a mess

& x is making a mess](Oscar). (Again, this list of determined propositions is not

exhaustive, but it will suffice for our purposes.)

Thus DRT is able to distinguish the truth conditions of Perry's utterances of (50)

and (51). Perry's utterance of (51) relative to DRS-8b is true only if Oscar holds the

attitude of belief toward the Russellian proposition denoted by '$\lambda x[x$ does not have a

leaky bag & x is not making a mess & x is making a mess](Oscar)'. The truth of Perry's

utterance of (50) relative to DRS-8b, however, does not require that Oscar hold the

attitude of belief toward such a contradictory proposition. (For the sake of simplicity, I

have slightly altered Perry's example. Perry's original example concerned the differing

truth conditions of occurrences of 'I believe that *he* is making a mess' and 'I believe that *I*

am making a mess'.)

I now turn to illustrating how DRT can be employed to debunk Richard's phone

booth argument. Richard initially employed his argument to support something akin to

Salmon's Russellian Theory, though he has since come to reject both that theory, and his

argument in support of it. Richard's argument is designed to show that even if, as the

Fregean theorists suggest, one does not permit "substitution within opaque contexts," the

semantic problems associated with the phenomenon of opacity will still arise. That is,

even if "substitution within opaque contexts" is banned, there will still be nomologically

possible situations in which occurrences of sentences of the form, $\ulcorner$N believes that $\Sigma(\alpha)\urcorner$

and $\ulcorner$N believes that $\Sigma(\beta)\urcorner$ (where α and β are coreferential, directly referring, terms) are

true, even though the subject referred to by N is disposed to assent to an occurrence of

$\Sigma(\alpha)$, yet is disposed to dissent from an occurrence of $\Sigma(\beta)$. One such nomologically

possible situation is as follows: Mark is wearing a disguise, and is in his room, which

contains a phone, and overlooks the street outside. Outside, across the street, there is a

phone booth which Mark can see from his window. The phone in Mark's room rings. Mark answers it and recognizes the voice of his old friend, Odile, and while talking to Odile, Mark looks outside and sees a woman whom he does not recognize at the phone booth. The woman at the phone booth is Odile, but because she is also wearing a disguise, Mark does not recognize her. Thus Mark does not realize that the person to whom he is speaking (Odile) is the same person whom he is watching. Suddenly Mark sees a steamroller heading toward the phone booth, and Mark thinks to himself (something like) "Gosh! The woman in the phone booth is in danger!" Our intuitions dictate that if Mark (for some reason) were to utter

(52) I believe that you are in danger.

into the phone, this utterance would be false. If, however, Mark were to utter

(53) I believe that she is in danger.

into the phone (where the occurrence of 'she' is accompanied by a gesture toward the woman in the phone booth), our intuitions dictate that Mark's utterance would be true. (Thus Richard would dissent from an occurrence of 'you are in danger' uttered into the phone, though he would assent to an occurrence of 'she is in danger', accompanied by a gesture toward the woman outside the window.) Richard's argument is designed to show that, despite our intuitions concerning the truth conditions of these occurrences, if Mark's utterance of (53) were true, then his utterance of (52) would also be true.

Suppose Mark tells Odile, over the phone, that he sees a woman at a phone booth, and that there is a steamroller heading toward this poor person. After describing what he sees to Odile, Mark utters an occurrence of,

(53) I believe that she is in danger.

into the phone. Our untutored intuitions dictate that this utterance is true. While telling Odile of the poor woman's plight, Mark waves his hands at the poor woman in an attempt to alert her of her danger. Odile sees Mark, whom she does not recognize, waving frantically and this leads Odile to tell Mark, over the phone, that there is a man who is waving frantically at her. She surmises--ironically enough--that the waving man believes that she is in danger. In telling Mark of this man who is waving at her and what she makes of it, Odile utters

(54) He believes that I am in danger.

into the phone (where the intended referent of 'He' is the man who is waving). After thinking about Odile's assessment of the waving man's behavior, Mark decides that Odile is correct. (Mark does not realize that the waving man is himself.) This leads Mark to voice his agreement, and thus he utters

(55) He believes that you are in danger.

into the phone (where again 'He' refers to the waving man). But notice that the only difference between (55) and (52) is that (55) has 'He' were (52) has 'I', and furthermore the relevant occurrences of these terms are coreferential, and they also appear *outside* of that-clauses--they appear in "transparent" positions. Hence, if Mark's utterance of (55) is true, then, despite our intuitions and despite the Fregean theorists ban of "substitution within opaque contexts," Mark's utterance of (52) must also be true. (Note the implicit appeal to the Principle of Propositional Truth!) The argument concludes that, since banning "substitution" does not succeed in preserving the veracity of our untutored

343

intuitions, we might as well abandon our intuitions and espouse some version of the Naive Russellian Theory.[2]

If Richard's argument is sound, then the veracity of our untutored intuitions must be rejected. I have claimed, however, that a Russellian theory can be made to accord with our untutored intuitions, and thus I must be able to show that Richard's argument is unsound. My objection to Richard's argument, in brief, is that it assumes the Principle of Propositional Truth; Richard's phone booth argument is unsound because it assumes the Principle of Propositional Truth, and this principle is false. Moreover, once the Principle of Propositional Truth is rejected, the above sketched machinery of DRT can be invoked to account for, and validate, our untutored intuitions concerning the truth conditions of occurrences of (52)-(55) in the above described situation.

Unlike other attempts to debunk Richard's phone booth argument, the DRT analysis of the truth conditions of the utterances of (52)-(55) crucially depends upon what other sentences are uttered in the discourse. Hence, before a DRT analysis can be provided, a more detailed description of Odile and Mark's discourse must be given. I assume that the following will suffice:

> Mark: There's a woman across the street, and (53) I believe that she is in danger.
>
> Odile: That the darndest thing, because there's a man in a window, and he's waving frantically. (54) He believes that I am in danger.
>
> Mark: Mmm. Yes. (55) He believes that you are in danger.

Let us assume that the DRS relative to which Mark's first string of utterances is produced is as follows:

<u>DRS-9a</u>

       Universe: *a, b*

       Conditions: 1. *a* is named 'Mark'
                  2. *b* is named 'Odile'
                  3. *a* and *b* are speaking with each other on the phone.

As the first sentence of Mark's utterance contains the indefinite NP, 'a woman', a new reference marker must be introduced into DRS-9a. The DRS which results from Mark's initial string of utterances, relative to DRS-9a is thus

<u>DRS-9b</u>

       Universe: *a, b, c*

       Conditions: 1. *a* is named 'Mark'
                  2. *b* is named 'Odile'
                  3. *a* and *b* are speaking with each other on the phone
                  4. *c* is across the street
                  5. *c* is a woman
                  6. *a* is in a belief state which is characterized by the following sub-DRS:

              Universe: *d*

              Conditions: 1. *d* is in danger

       Internal Restrictions: 1. $c \leftrightarrow d$

       External Restrictions: 1. R(*a*) = Mark
                            2. R(*d*) = Odile

Note that the DRS is constructed in two steps: In the first step, the reference marker *c*, and conditions 3 and 4 are introduced into the primary DRS, and in the second step condition 5 is introduced, as well as internal restriction 1. And finally, as 'she' is--even

345

when it appears inside a that-clause--a directly referential term, the external restriction 2 is added. (As usual, I have suppressed any details which are not directly relevant.)

Odile's response(s) to Mark's initial string of utterances brings about the following DRS:

DRS-9c

> Universe: *a, b, c, e*
>
> Conditions: 1. *a* is named 'Mark'
> 2. *b* is named 'Odile'
> 3. *a* and *b* are speaking with each other on the phone
> 4. *c* is across the street
> 5 *c* is a woman
> 6. *a* is in a belief state which is characterized by the following sub-DRS:
>
>>   Universe: *d*
>>
>>   Conditions: 1. *d* is in danger
>
> 7. *e* is at a window, waving frantically
> 8. *e* is a man
> 9. *e* is in a belief state which is characterized by the following sub-DRS:
>
>>   Universe: *f*
>>
>>   Conditions: 1. *f* is in danger
>
> Internal Restrictions: 1. *c* ↔ *d*
> 2. *f* ↔ *b*
>
> External Restrictions: 1. R(*a*) = Mark
> 2. R(*d*) = Odile
> 3. R(*e*) = Mark
> 4. R(*f*) = Odile

Again, DRS-9c is brought about in two steps: First, the indefinite NP 'a man in a window' introduces a new reference marker, viz. *e*, into the universe, and it also

introduces conditions 7 and 8. Second, the ascription uttered in Odile's response(s) adds condition 9, and the internal restriction, $f \leftrightarrow b$. Also, as 'I' is a directly referential term, the external restriction $R(f) = $ Odile is added to the DRS.

Mark's last string of utterances merely reiterates the information conveyed by Odile's previous utterance, and thus it induces no significant changes in DRS-9c.

Our intuition dictate that in the described situation all of utterances (53)-(55) are true. Though I will not go through the details, the above sketched DRT analysis preserves these intuitions. But what about (52)? What would be the truth value of Mark's utterance of

(52) I believe that you are in danger.

relative to DRS-9c? I suggest that this occurrence would induce the following DRS:

<u>DRS-9d</u>

Universe: *a, b, c, e*

Conditions: 1. *a* is named 'Mark'
     2. *b* is named 'Odile'
     3. *a* and *b* are speaking with each other on the phone
     4. *c* is across the street
     5. *c* is a woman
     6. *a* is in a belief state which is
      characterized by the following sub-DRS:

    Universe: *d, g, h*

    Conditions: 1. *d* is in danger
         2. *g* is in danger
         3. *g* is named 'Odile'
         4. *h* and *g* are speaking with each other on the phone
         5. *g* is the referent of Mark's utterance of 'you'

    7. *e* is at a window, waving frantically
    8. *e* is a man
    9. *e* is in a belief state which is
     characterized by the following sub-DRS:

   Universe: *f*

   Conditions: 1. *f* is in danger

Internal Restrictions: 1. *c* ↔ *d*
        2. *f* ↔ *b*
        3. *g* ↔ *b*
        4. *h* ↔ *a*

External Restrictions: 1. R(*a*) = Mark
         2. R(*d*) = Odile
         3. R(*e*) = Mark
         4. R(*f*) = Odile
         5. R(*g*) = Odile

Mark's utterance of (52) contains an occurrence of the definite NP 'you' and this definite NP must correspond with a reference marker which is linked to a reference marker already present in DRS-9d. The only plausible candidate is reference marker b. Hence the reference marker corresponding with the occurrence of 'you' in Mark's utterance of (52), viz. reference marker g, is linked via internal restriction 3 to reference marker b. But why are conditions 3-5 added to the sub-DRS of condition 6? Mark's utterance of (52) is a *self-ascription*; thus the subject of the ascription, Mark, is also *a participant* of the discourse. Hence if Mark were to utter an occurrence of (52), he would understand the NPs of the discourse as standing in the coreferring relationships represented by the internal restrictions of DRS-9d, and thus if Mark were to sincerely utter (52), he would assume that the token of 'you' appearing in his utterance of (52) refers to a person named 'Odile' with whom he is talking on the phone. This is represented in DRT as follows: the sub-DRS in condition 6 characterizes Mark's belief state, yet DRS-9d itself--which contains the internal restriction $g \leftrightarrow b$--also characterizes Mark's belief state. Therefore Mark's utterance of the self-ascription (52) would have the effect of *copying* the conditions in DRS-9c which govern reference marker b into the sub-DRS of condition 6. (Though these *copied conditions* govern reference markers g and h, instead of internally linked reference markers a and b.) Thus conditions 2-3 from DRS-9c are *copied* as conditions 3 and 4 to the sub-DRS of condition 6 in DRS-9d. I assume that conditions 2 and 5 of the sub-DRS in condition 6 require no explanation.[3]

The above DRT analysis of Mark's supposed utterance of (52) explains a rather odd feature of Richard's argument: Richard's argument asks us to consider what the truth conditions of Mark's utterance of (52) would be, though the details of the described situation make it clear that Mark *would not* sincerely utter (52); Mark would judge his own utterance of (52) to be false. A proper analysis of Mark's possible utterance of (52) ought to explain why Mark would not sincerely utter (52), and the above DRT analysis provides such an explanation: Mark would not sincerely utter (52) because, relative to

DRS-9c, an occurrence of (52) uttered by Mark communicates that he believes that the person he is speaking with on the phone--the referent of his utterance of 'you'--is in danger, and Mark believes this to be false information.

What about truth conditions? Richard's argument in support of the Naive Russellian theory goes through only if it follows from the truth of Mark's utterance of (55) that Mark's subsequent utterance of (52) is also true. That is, Richard's argument goes through only if it cannot be that Mark's utterance of (55) relative to DRS-9c is true, while Mark's supposed utterance of (52) relative to DRS-9c is false. Under our definition of truth for an occurrence, however, it is possible for the former utterance to be true, and the latter false. Under our definition for truth for an occurrence, Mark's utterance of (52) relative to DRS-9c is true only if Mark holds the attitude of belief toward, among other things, the Russellian proposition that $\lambda x[\lambda y[x$ is in danger $\& x$ is named 'Odile' $\& y$ and x are speaking with each other on the phone](Mark)](Odile). The truth of Mark's utterance of (54), however, does require that Mark hold the attitude of belief toward this proposition. Thus the truth conditions of the occurrences differ, and Richard's argument in support of the Naive Russellian Theory is debunked.

Appendix B Notes

[1] Perry presents his problem in Perry 1979, and Richard presents his phone booth argument in Richard 1983. In Asher 1986, a version of DRT is utilized in order to debunk Richard's argument, though the analysis presented here differs from Asher's in several respects.

[2] In Soames 1987a, Soames argues that Richard's argument applies in similar situations which involve occurrences of proper names instead of indexicals.

[3] It appears as if condition 5 of the sub-DRS of condition 6 contains a kind of use-mention confusion. This condition states

> 5. g is the referent of Mark's utterance of 'you'

but 'g' refers to a *reference marker* and not to Odile. Recall, however, that for the sake of simplicity the predicates appearing the conditions of DRSs serve "double duty"; the predicate 'is a dog' appearing in a condition does not, strictly speaking, designate the property *being a dog*, but instead designates a mental representation type, tokens of which are the mental representation tokens which allow the participants of the relevant discourse to seemingly speak and think about the same property, the property *being a dog*. Thus, strictly speaking, the predicate appearing in condition 5 designates a certain mental representation type, tokens of which allow Mark and Odile to speak and think about the same property, the property of being the referent of Mark's utterance of 'you'.

Bibliography

Asher, N. 1986. "Belief in Discourse Representation Theory." *Journal of Philosophical Logic* 15:137-189.

Austin. J.L. 1962. *How to do things with Words.* Oxford: Oxford University Press.

Barwise. J. and Perry, J. 1983. *Situations and Attitudes.* Cambridge: MIT Press.

Berg. 1988. "The Pragmatics of Substitutivity." *Linguistics and Philosophy* 11:355-370.

Block, N. 1980. *Readings in the Philosophy of Psychology, Vol. 1.* Cambridge: Harvard University Press.

Block, N. 1986. "Advertisement for Semantics for Psychology." *Midwest Studies in Philosophy, X.*

Boer and Lycan. 1985. *Knowing Who.* Cambridge: MIT Press.

Boghossian, P. 1989. "Content and Self-Knowledge." *Philosophical Topics*, vol. XVII, spring.

Braithwaite, R.B. (ed.). 1931. *The Foundations of Mathematics.* London: Routledge.

Burge, T. 1977a. "Kaplan, Quine and Suspended Belief." *Philosophical Studies* 31:197-203.

Burge, T. 1977b. "Belief *de Re.*" *Journal of Philosophy* 69:338-362.

Burge, T. 1979 :Indivualism and the Mental." In P. French, T. Uhling and H. Wettstein (eds.) *Midwest Studies in Philosophy 4.* Minneapolis: University of Minnesota Press.

Burge, T. 1988. "Individuation and Self-Knowledge." *Journal of Philosophy*, 85.1 (Nov.).

Cartwright, R. 1987. *Philosophical Essays.* Cambridge: The MIT Press.

Castaneda, H-N. 1968. "On the Logic of Attributions of Self-Knowledge to Others." *Journal of Philosophy* 65:439-456.

Church, A. 1951. "A Formulation of the Logic of Sense and Denotation." In Henle, P., Kallen, H., and Langer, S., (eds.). *Structure, Method and Meaning: Essays in Honor of H.M. Sheffer.* New York: Liberal Arts Press.

Crimmins, M. 1992a. *Talk About Beliefs.* Cambridge: MIT Press.

Crimmins, M. 1992b. "Context in the Attitudes." *Linguistics and Philosophy* 15:185-98.

Crimmins, M. and Perry, J. 1989. "The Prince and the Phone Booth: Reporting Puzzling Beliefs." *The Journal of Philosophy* 86:685-711.

Davidson, D. 1967. "Truth and Meaning." In *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press, 1990.

Davidson, D. 1968. "On Saying That." In *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press, 1990.

Davidson, D. 1984. "First Person Authority." *Dialectical,* XXXVIII, 2-3: 101-111.

Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association,* LX: 448

Davidson, D. 1988. "Reply to Burge." *Journal of Philosophy,* 85.1 (Nov.).

Davidson, D. 1990. *Inquiries into Truth and Interpretation.* Oxford: Oxford University Press.

Dennett, D. 1971. "Intentional Systems." *Journal of Philosophy* LLXVIII, 4:87-106.

Donnellan, K. 1966. "Reference and Definite Descriptions." *Philosophical Review* 75:647-658.

Dretske, F. 1981. *Knowledge and the Flow of Information.* Cambridge: MIT Press.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes.* Cambridge: MIT Press.

Dummett, M. 1978. *Truth and Other Enigmas.* Cambridge: Harvard University Press.

Dummett, M. 1973. *Frege: Philosophy of Language.* Cambridge: Harvard University Press.

Evans, G. 1977. "The Causal Theory of Names." In *Naming, Necessity, and Natural kinds.* Ithaca: Cornell University Press.

Evans, G. 1981. "Understanding Demonstratives." In H. Parret and J. Bouveresse (eds.), *Meaning and Understanding,* Berlin: de Gruyter, 1981.

Evans, G. 1982. *The Varieties of Reference.* Oxford: 1982.

Fodor, J. 1981. "Propositional Attitudes." In *Representations: Philosophical Essays on the Foundations of Cognitive Science* Cambridge: MIT Press, 1981.

Fodor, J. 1990. *A Theory of Content and Other Essays.* Cambridge: MIT Press.

Forbes, G. 1990. "The Indispensability of Sinn." *Philosophical Review,* 1990, 99:535-564.

Frege, G. 1966. P. Geach and M. Black, (eds.). *Translations from the Philosophical Writings of Gottlob Frege.* London: Basil Blackwell.

Frege, G. 1979. Hermes, Kambartel and Kaulbach (eds.). *Posthumous Writings: Gottlob Frege.* Chicago: University of Chicago Press.

Frege, G. 1988. "The Thought." In N. Salmon and S. Soames (eds.) *Propositions and Attitudes* Oxford: Oxford University Press.

Gabrial, Hermes, Kambartel, Theil and Veraart (eds.). 1980. *Frege, Philosophical and Mathematical Correspondence.* Translated into English by Hans Kaal. Oxford: Blackwell.

Geach, P. 1962. *Reference and Generality.* Ithaca: Cornell University Press.

Goldman, A. 1986. *Epistemology and Cognition.* Cambridge: Harvard University Press.

Grice, H. 1957. "Meaning." *Philosophical Review,* 66:147-177.

Grice, H. 1968. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning." *Foundations of Langugage,* 4:225-242.

Grice, H. 1975. "The Logic of Conversation." In D. Davidson and G. Harman (eds.), *The Logic of Grammar.* Encino: Dickenson.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases,* (University of Massachusetts, Amherst Ph.D. dissertation.

Heim, I. 1983. "File Change Semantics and the Familiarity Theory of Definiteness." In Bauerle *et al.* (eds.) *Meaning, Use and Interpretation of Language* (Berlin: 1983).

Higginbotham, J. 1991. "Belief and Logical Form." *Mind and Language,* Vol. 6, No. 4, Winter.

Hintikka, J. 1980. "On Sense and Reference and the Objects of Knowledge. *Epistemolgia* vol. 3, pp. 143-162.

Kamp, H. 1981. "A Theory of Truth and Semantic Representation." In Groenendijk, Janssen and Stokhof (eds.) *Formal Methods in the Study of Language* (Amsterdam: 1981).

Kamp, H. 985. "Context, Thought and Communication." *Proceedings and Addresses of the Aristotelian Society*, 85.

Kamp, H. 1988 "Comments on Stalnaker." In R. Rimm and D. Merrill (eds.), *Contents of Thought*, 156-181. Tuscon: University of Arizona Press.

Kamp, H. 1990. "Prolegomena to a Structural Account of Belief and Other Attitudes." In C.A. Anderson and J. Owens (eds.), *Propositionsal Attitudes*, 27-90. Stanford, CA: CSLI..

Kaplan, D. 1969. "Quantifying In." *Synthese* 19:178-214.

Kaplan, D. 1975. "How to Russell a Frege Church." *Journal of Philosophy* 72: 716-729.

Kaplan, D. 1978 "Dthat." In P. Cole (ed.) *Syntax and Semantics 9: Pragmatics*, 221-243. New York: Academic Press.

Kaplan, D. 1989a. "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Kaplan, D. 1989b. "Afterthoughts." In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Katz, J. 1990. *The Metaphysics of Meaning*. Cambridge: MIT Press.

Kripke, S. 1972. *Naming and Necessity*. Cambridge: Harvard University Press.

Kripke, S. 1977. "Speaker's Reference and Semantic Reference" In P. French, T. Uhling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press.

Kripke, S. 1979. "A Puzzle About Belief." In Margalit A. (ed.) *Meaning and Use*. Dordrecht: D. Reidel.

Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lewis, D. 1979a. "Scorekeeping in the Language Game." *Journal of Philosophical Language* 8:339-359.

Lewis, D. 1979b. "Attitudes *De Dicto* and *De Se*." *The Philosophical Review*, 88.

Lewis, D. 1981. "What Puzzling Pierre Believes." Australian Journal of Philosophy 59, no. 3:283-289.

Lewis, D. 1983. *Philosophical Papers, Vol. 1*, Oxford: Oxford University Press.

Marcus, R. B. 1990. "Some Revisionary Proposals About Belief and Believing." *Philosophy and Phenomenological Research*, Vol. L, Supplement, Fall.

Mates, B. 1950. "Synonymity." *University of California Publications in Philosophy* 25:201-226.

Perry, J. 1977. "Frege on Demonstratives." *Philosophical Review* 86:474-497.

Perry, J. 1979. "The Problem of the Essential Indexical." *Nous* 13:3-21.

Perry, J. 1980. "Belief and Acceptance." In P. French, T. Uehling, and H. Wettstein (eds.) *Midwest Studies in Philosophy*, Vol. 5, 533-542. Minniapolos: University of Minnesota Press.

Perry, J. 1986. "Thought without Representation." *Proceedings of the Aristotelian Society*, suppl. vol. LX, pp. 263-83.

Perry, J. 1988. "Cognitive Significance and New Theories of Reference." *Nous* 22:1-18.

Putnam, H. 1975. "The Meaning of 'Meaning'." In *Mind, Language and Reality: Philosophical Papers*. Cambridge: Cambridge University Press.

Quine, W.V. 1953. Reference and Modality." In *From a Logical Point of View*. Cambridge: Harvard University Press.

Quine, W.V. 1960. *Word and Object*. Cambridge: MIT Press.

Quine, W.V. 1966. "Quantifiers and Propositional Attitudes." In *The Ways of Paradox*. New York: Random Houuse.

Richard, M. 1983. "Direct Reference and Ascriptions of Belief." *Journal of Philosophical Logic* 12:425-452.

Richard, M. 1990. *Propositional Attitudes: An Essay on Thoughts and How we Ascribe Them*. Cambridge: Cambridge University Press.

Richard, M. 1992. "Attitudes in Context." *Linguistics and Philosophy*, 16, pp. 123-48.

Russell, B. 1903. *The Principles of Mathematics*. London: Norton.

Russell, B. 1956a. "On Denoting." In R. March (ed.), *Logic and Knowledge*, 39-56. London: Unwin Hyman.

Russell, B. 1956b. "On the Nature of Acquaintance." In R. March (ed.), *Logic and Knowledge*, 125-174. London: Unwin Hyman.

Russell, B. 1959. *The Problems of Philosophy*. Oxford: Oxford University Press.

Salmon, N. 1986. *Frege's Puzzle*. Cambridge: MIT Press.

Salmon, N. 1989. "Illogical Belief." *Philosophical Perspectives, 3 Philosophy of Mind and Action Theory*, pp. 243-85.

Salmon, N. and Soames, S., (eds.). 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.

Scheffler, I. 1955. "On Synonomy and Indirect Discourse." In *Inquiries: Philosophical Studies of Language, Science and Learning*. Indianapolis: Hacket, 1986.

Schiffer, S. 1989. *Remnants of Meaning*. Cambridge: MIT Press.

Schiffer, S. 1990. "The Mode of Presentation Problem." In Anderson, C. and Owens J. (eds.) *Propositional Attitudes*. Stanford: CSLI, 1990.

Schiffer, S. 1992. "Belief Ascription." *The Journal of Philosophy vol. LXXXIX*, no. 10.

Soames, S. 1985. "Lost Innocence." *Linguistics and Philosophy* 8:59-71.

Soames, S. 1987a. "Direct Reference, Propositional Attitutudes, and Semantic Content." Philosophical Topics." 15:44-87.

Soames, S. 1987b. "Substitutivity," in JJ Thomson (ed.) *On Being and Saying: Essays in Honor of Richard Cartwright*. Cambridge: MIT Press.

Sperber, D. and Wilson, D. 1988. *Relevance: Communication and Cognition*. Cambridge: Harvard University Press

Stalnaker, R. 1974. "Pragmatic Presupposition." In M.K. Muntz and P.K. Unger (eds.) *Semantics and Philosophy*, New York: New York University Press.

Stalnaker, R. 1979. "Assertion." In Cole (ed.), *Syntax and Semantics 9, Pragmatics*, New York: Academic Press.

Stalnaker, R. 1981. "Indexical Belief." *Synthese* 49.

Stalnaker, R. 1987a. *Inquiry*. Cambridge: The MIT Press.

Stalnaker, R. 1987b. "Semantics for Belief." *Philosophical Topics*, Vol. XV. No. 1 Spring.

Stalnaker, R. 1988. "Belief Attribution and Context." In R. Rimm and D. Merrill (eds.), *Contents of Thought*, 156-181. Tuscon: University of Arizona Press.

Stalnaker, R. 1990. "Narrow Content." In *Propositional Attitudes: The Role of Content in Logic, Language and Mind.* C. Anthony Anderson and Joseph Owens, (eds.) CSLI Lecture Notes, no. 20.

Stich, S. *From Folk Psychology to Cognitive Science: The Case Against Belief.* Cambridge: MIT Press.

Wettstein, H. 1991. *Has Semantics rested on A Mistake? And other Essays.* Stanford: Stanford University press.