Utrecht University

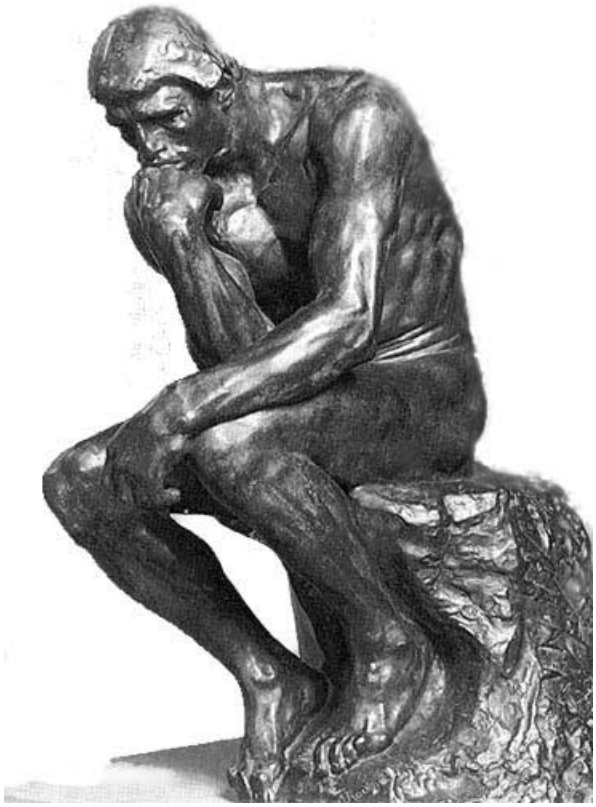
Bachelor's Thesis for Artificial Intelligence

To think or not to think?

The use of language in a theory of consciousness

Author:
Renata Fonville

Supervisor:
Dr. Menno Lievers



July 1, 2010

Contents

1 Introduction

In studying artificial intelligence, the study of applying our knowledge of human (or animal) beings onto artificial intelligent systems, we stumble from time to time upon the question of whether there will come a time that those systems are going to be *just like us*. Of course this question does not pertain to their intelligence; we already know we can make artificial systems at least as intelligent as we are. No, the question rather concerns everything which we so easily identify as "what makes us human", the focus being on 'consciousness'. In short, researchers in artificial intelligence wonder whether some day artificial intelligent systems, like robots for example, can be conscious. In order to find this out we need to define what 'consciousness' is. This is easier said than done, and so we end up in the ongoing philosophical debate about the mystery of consciousness.

What is consciousness? What does consciousness consist of? How does consciousness arise? We can ask a lot of questions about the phenomenon we call 'consciousness', but to this day there still is no answer upon which we all agree. Theories about consciousness are all over the place: some are determined to find the solution in the brain, while others believe the riddle has to be solved conceptually. While some like to believe consciousness can be present in animals, others boldly say that consciousness is reserved only for us human beings.

When talking about consciousness, there are two distinct kinds of things that could be the subjects we ascribe this property to. We could be talking about a creature being awake and responsive ("He has regained his consciousness!"). Or we could be talking about the mental states such a creature could be in, such as thoughts, desires, emotions, and sensations, as being conscious or not conscious. The property of consciousness for a creature is distinct from the property of consciousness for a mental state. In what follows it is the latter notion of consciousness, which we could call *state consciousness*, that we shall be concerned with.

1.1 A higher-order theory of consciousness

Higher-order theories of consciousness seek to explain this state consciousness, concerning themselves primarily with the difference between conscious states and states that are not conscious. Well-known thinkers like Aristotle, Descartes, Locke and Kant all held that we are in some way conscious of our conscious states. Higher-order theories embrace this idea as well: a subject must be appropriately conscious of a mental state in order for it to be a conscious mental state. It is because being conscious of a state involves some higher-order awareness that we call theories that adopt this idea higher-order theories.

The view that a state's being conscious consists in one's being conscious of that state can also be referred to as the *transitivity principle*[1]. The transitivity principle does not by itself specify the way in which one must be conscious of a state in order for that state to be conscious. So even though higher-order theories of consciousness all endorse the transitivity principle, they differ a lot in their views of how the principle is implemented. This means that today there exist several different higher-order theories of consciousness.

[1] Rosenthal 1997

The philosopher David Rosenthal has presented an interesting higher-order theory of consciousness: the higher-order thought (HOT) theory. This theory tries to explain state consciousness in terms of higher-order thoughts and relies, when implementing that idea, very much on how we use language. In this thesis we will take a close look at Rosenthal's HOT theory and its implementation. While doing that we will see that even though the HOT theory is nicely established and very well thought through, it still faces some difficulties.

1.2 The use of language in a theory of consciousness

We will see that those difficulties mainly derive from the fact that the HOT theory in its current form is thoroughly based on the use of language. When Rosenthal is explaining and verifying the fundamental concept of the HOT theory he is relying solely on the use of language while not being consistent with how we actually use language, which makes the plausibility of the basic concept of the HOT theory questionable.

Explaining a theory of consciousness, especially one like the HOT theory, through how we use language is not the right way to go. Language and, moreover, how we use it is very multi-interpretable and both language itself and how we interpret it are continually subject to change. Those attributes of language and its use make it a very unsuitable way of supporting a theory of consciousness.

To show why the use of language is such an unsuitable way to explain a theory of consciousness we will first, in section 2 (*Rosenthal's HOT theory*), make ourselves familiar with the HOT theory as Rosenthal has presented it. Then, in section 3 (*Problems with the use of language in the HOT theory*), we will explore the problems this theory encounters caused by its reliance on the use of language. These problems will show to be too significant to ignore, which, in section 4 (*Consequences for the HOT theory*) will give us the opportunity to conclude that a theory of consciousness should not have language and its use as its main support.

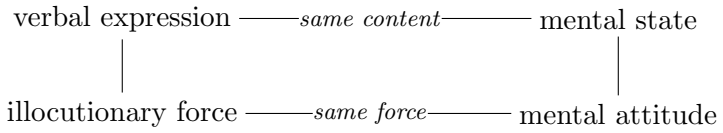
2 Rosenthal's HOT theory

It is said that we all have mental states; some of them are conscious, some of them are not. What is the difference between conscious mental states and unconscious mental states? What is it that makes a mental state conscious? David Rosenthal thinks he has an answer to this: he has developed a higher-order thought theory of consciousness. For the theory to work, Rosenthal assumes that in general, being conscious or aware of something is to have a thought about it. You may be conscious of this thesis, for example, by having some thought about it. Similarly for mental states: a mental state of yours may be conscious if you have a suitable thought about it. From there on the theory rests on the claim that "a mental state's being conscious is its being accompanied by a roughly simultaneous higher-order thought about that very mental state"[2]. On this account, what it is that makes a mental state conscious is a higher-order thought (HOT), and conscious mental states differ from those which are not in them being accompanied by a suitable HOT.

2.1 Verbally expressing and reporting

Rosenthal himself emphasizes the importance of the distinction between verbally expressing a thought and reporting a thought for his HOT theory. The distinction here, according to Rosenthal, is that there are two ways of conveying one's thought that p: either just by saying that p or by saying that one thinks that p. Then, by saying that p you are *verbally expressing* your thought that p, and by saying that you think that p you are *reporting* your thought that p. When you have a thought that it's raining, for example, you can verbally express that thought by saying "It's raining", or you can report that thought by saying "I think it's raining".

The difference between verbal expressions and reports can also be made clear when looking at correspondences in content or force. When you think it's raining and you verbally express that thought by saying that it's raining, your verbal expression has the same content as the mental state it expresses, namely that it's raining. The verbal expression, or speech act, also has an illocutionary force that corresponds to the mental attitude of the mental state it expresses. This means that a verbal expression and the mental state it expresses have the same force, which could be that of suspecting, denying or wondering for example. This relation between a verbal expression and the mental state it expresses is made visible in the diagram below.



Now by contrast, when you think it's raining and you report that thought by explicitly saying that you think it's raining, your report does obviously not have the same content as the mental state it reports. Further, if we stick to our example about the thought that it's raining, the report will still have an illocutionary force that corresponds to the mental

[2] Rosenthal 2005b

attitude of the mental state it reports: the illocutionary force of "I think it's raining" and the mental attitude of a thought that it's raining are both that of asserting. But the contrast would be more decisive if we take another example with a nonassertoric mental attitude. If you wonder whether it'll rain, for example, you can verbally express that state of wondering by saying "Will it rain?", or you can report that state by saying "I wonder whether it'll rain". Here the illocutionary force of the verbal expression still corresponds to the mental attitude of wondering, but the illocutionary force of the report, on the other hand, does not correspond to the mental attitude of wondering. The illocutionary force of the report is that of asserting, as it is with all reports of intentional states.

Rosenthal insists that the phenomenon we know as Moore's paradox is helpful in recognizing the distinction between verbally expressing and reporting. G.E. Moore famously observed that sentences such as "It's raining but I don't think it is" cannot be used to make coherent assertions, even though they are not contradictory.[3] Nowadays, sentences of the form 'p but I don't think that p' have widely become known as Moore's paradox. The reason for not being able to assert anything of the form of Moore's paradox is because the assertion of the first conjunct ("It's raining") would express an intentional state that the second conjunct ("but I don't think it is") denies I am in. Rosenthal says that if there were no difference between verbally expressing an intentional state and reporting it

> then my denial that I am in the intentional state of thinking that it's raining would be tantamount simply to expressing the thought that it's not raining. Accordingly, the Moore's-paradox sentence would be equivalent to "It's raining and it's not raining", which is an actual contradiction. [...] To avoid this result, we must distinguish reporting our intentional states from verbally expressing them.[4]

An important distinction between verbal expressions and reports is that they have different semantic properties: they mean different things and have distinct truth conditions. But aside from that, the distinction between verbally expressing and reporting one's mental states is, according to Rosenthal, a distinction that is not appreciated very easily. This could be the case due to the fact that verbal expressions and reports usually have the same performance conditions: whenever it is appropriate to say "It's raining" it would also have been appropriate to say "I think it's raining", which amounts to an easy conflation of the concepts of verbal expressions and reports. But Rosenthal does not want to permit this kind of conflation:

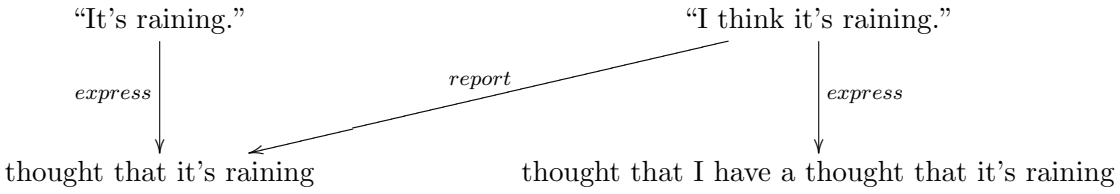
> If saying that p amounted to the same thing as saying that one thinks that p, expressing a state would be the same as expressing one's consciousness of that state, which would encourage identifying mental states with one's consciousness of them.[5]

2.2 HOT theory in action

Now how does the HOT theory of Rosenthal explain consciousness with the help of the distinction between verbally expressing and reporting? Consider the utterance "It's raining". This utterance verbally expresses the thought someone has that it's raining. Now consider the

[3] Moore 1942
[4] Rosenthal 2005a
[5] Rosenthal 2005c

utterance "I think it's raining". This utterance reports the thought that it's raining, and also verbally expresses the thought that I have a thought that it's raining, which we can also call the HOT about the thought that it's raining. So when we report our thought that it's raining we are simultaneously verbally expressing our HOT about the thought that it's raining, and since we have that HOT about the thought that it's raining, the thought that it's raining is a conscious one. This relation between first-order thoughts and HOTs, and reporting and verbally expressing them, is made visible in the diagram below.



From this equality between reporting a mental state and verbally expressing the HOT about that mental state, Rosenthal infers that we can report a mental state we are in just in case that state is conscious, and vice versa that when a mental state is not conscious we cannot report being in it[6]. In short we can say that if we have a HOT about a mental state, we can report that mental state; and if we can report a mental state, that mental state is conscious.

HOTs rarely are conscious, which is not very strange, since for a HOT to be conscious there must be an even *higher*-order thought about that HOT. But in the case there *is* such a higher-order thought, we can speak of introspective consciousness. Being introspectively conscious of a mental state means that you are aware of being conscious of that mental state.

2.3 Why verbally expressed thoughts are conscious

So far we have stated that Rosenthal's HOT theory of consciousness predicts that if we can report a mental state, that mental state is conscious. But next to the ability to *report* mental states as an indication of them being conscious, Rosenthal also claims that *verbally expressing* mental states indicates that they are conscious. He puts forward that though reporting a mental state and verbally expressing that mental state are semantically distinct (they mean different things and have distinct truth conditions), they are performance conditionally equivalent. This means that whenever it is appropriate for me to say that it's raining, it is also appropriate for me to say that I think it's raining.

According to Rosenthal this performance-conditional equivalence is second nature for us. By adding this second-nature character Rosenthal wants to imply that whenever one says "It's raining", thereby verbally expressing a mental state, one could just as easily have said "I think it's raining" and thereby reporting that mental state. Rosenthal concludes that not only *reported* thoughts are conscious, but also that *verbally expressed* thoughts are in general conscious. In his own words:

> One's verbally expressed thoughts are in general conscious because the speech acts that express them have the same performance conditions as reports about those

[6] Rosenthal 2005c

thoughts, and that equivalence is second nature for us.[7]

However, when it comes to verbally expressing a HOT, instead of a first-order thought, Rosenthal says that

> when we verbally express a HOT, that verbal expression is performance conditionally equivalent to a report of that HOT, but the HOT still need not be conscious because that performance-conditional equivalence is not second nature for us.[8]

To illustrate this we can once again use the example about the rain. From the performance-conditional equivalence between verbal expressions and reports of HOTs follows that whenever it is appropriate for me to say that I think it's raining, it is also appropriate for me to say that I think that I think it's raining. But, following Rosenthal, this performance-conditional equivalence is not second nature for us, meaning that whenever I say "I think it's raining" I could *not* just as easily have said "I think that I think it's raining".

Thus verbally expressed thoughts are conscious since verbally expressing them and reporting them are performance conditionally equivalent, which is second nature for us, and verbally expressed HOTs need not be conscious since the performance-conditional equivalence between verbally expressing HOTs and reporting HOTs is not second nature for us.

[7] Rosenthal 2005c
[8] Rosenthal 2005c

3 Problems with the use of language in the HOT theory

The way Rosenthal has built up his HOT theory of consciousness is very sophisticated. All the parts link together very well, and if everything in his theory were entirely correct the HOT theory would actually be a very good candidate for explaining consciousness satisfactorily. However, even though Rosenthal's HOT theory of consciousness thus sounds very plausible, you might have noticed that when you try its examples for yourself there is something more going on than what Rosenthal is telling us. When you say "I think it's raining", are you conscious of a certain thought you have that it is raining and trying to report that thought? No, that is probably not the case. As it turns out, there is more to the verb 'to think' than just literally 'having thoughts'.

As mentioned in the introduction, this is a problem that arises from the great emphasis that is placed on language in the HOT theory, and in particular on how we use language. A large part of the theory relies on and is explained by means of making utterances: expressing thoughts and reporting thoughts. When there is such a reliance on the use of language it should be in accordance with how we actually use language. But this is not the case in the HOT theory and the next couple of sections will explain why.

3.1 Interpretation

After learning of Rosenthal's HOT theory we can think of three messages you are intending to get across when uttering the sentence "I think it's raining":

(1) you have a thought that it's raining

(2) it's raining

(3) you think it's raining, but you're not sure

The messages in (1) and (2) are coming from the usage of the utterance in the HOT theory and what Rosenthal taught us about reports and expressions. The message in (3), on the other hand, is a new one. However, it sounds very familiar, which is exactly the point.

In (1) we find the meaning of 'to think' of which Rosenthal explicitly tells us it is how he wants us to interpret this verb: 'to think' means 'to have mental states'. We immediately encounter a problem here: when saying "I think it's raining", do you ever literally mean to say that you have a thought that it's raining? No. The message in (1), the one Rosenthal needs to be the case for his HOT theory, is not a common interpretation of the utterance we are considering. Similarly, the message in (2) is not a common interpretation of the utterance "I think it's raining" either.

The message in (3) is one Rosenthal does not want us to consider when looking at the HOT theory. For his current theory this is a smart choice, since we will see that when we do seriously consider that message the HOT theory of consciousness fails at certain points. Unfortunately for Rosenthal, this familiar-sounding message we find in (3) is in fact a common interpretation of the utterance we are considering, which means that we cannot just ignore it.

To get a grasp of why the messages in (1) and (2) are not common interpretations of the utterance "I think it's raining" and why the message in (3) is, we will have to take a look in the field of pragmatics; in particular at the theory of conversational implicatures.

3.2 Conversational implicatures

The notion of conversational implicatures was initially coined by Paul Grice[9] within his theory about how people use language and behave in conversation. In this theory Grice suggests that there exists a set of over-arching assumptions guiding the conduct of conversation, which arise from rational considerations. Grice identifies them as guidelines for the efficient and effective co-operative use of language:

1. *The maxim of Quality*
   Try to make your contribution one that is true, specifically:
   (a) do not say what you believe to be false
   (b) do not say that for which you lack adequate evidence
2. *The maxim of Quantity*
   (a) make your contribution as informative as is required for the current purposes of the exchange
   (b) do not make your contribution more informative than is required
3. *The maxim of Relevance*
   Make your contribution relevant
4. *The maxim of Manner*
   Be clear, and specifically:
   (a) avoid obscurity
   (b) avoid ambiguity
   (c) be brief
   (d) be orderly

In short, these guidelines, or maxims, specify that participants in a conversation should speak sincerely, relevantly and clearly, while providing sufficient information. Grice's maxims of conversation altogether underlie the co-operative principle:

> make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.[10]

Although the maxims of conversation and the co-operative principle seem phrased as prescriptive commands, they are intended as a description of how people normally behave in conversation. And this behavior goes both ways: speakers generally observe the co-operative principle and its maxims and listeners generally assume that speaker are indeed observing

[9] Grice 1975
[10] Grice 1975

it. We do however have to understand that Grice's point is not that people follow these guidelines to the letter, but still that people are generally being co-operative at some deeper level. Now this is where conversational implicatures can arise. Consider the following short dialogue:

(4) A: Where's Bill?
B: There's a yellow VW parked outside Sue's house.

Obviously, when we take B's contribution literally it fails to answer A's question and seems to violate at least the maxims of Quantity and Relevance, therefore not being a co-operative contribution. However, it seems apparent that we nevertheless try to interpret B's contribution as co-operative at some deeper level, by assuming that it is in fact co-operative. Then we can infer that Bill owns a yellow VW, which means he might be in Sue's house.

We can see that in cases like this, inferences arise to preserve the assumption of co-operation. It is this kind of inferences that Grice calls conversational implicatures.

3.3 HOT theory vs. conversational implicatures

We have noted that a common interpretation of saying that I think it is raining is that I think it might be raining, but I am not completely sure about that. In other words: when saying "I think it's raining", the implicature is that I do not know for a fact that it is raining. This implicature of that utterance and the whole notion of conversational implicatures is completely ignored by Rosenthal, as the following quote will typically illustrate:

> If I think that the door is open, for example, I can convey this thought to you simply by saying 'The door is open'; that speech act will express my thought. But I could equally well convey the very same thought by saying, instead, 'I think the door is open'. Similarly, I can communicate my suspicion that the door is open either by expressing my suspicion or by explicitly telling you about it. I express the suspicion simply by saying, for example, that the door may well be open, whereas I would be explicitly telling you that I have that suspicion if I said that I suspect that it is open.[11]

As we can see, the suspicion Rosenthal mentions in his second example already arises by means of an implicature in his first example, which indicates that both of Rosenthal's examples are actually not that different. The following two examples of different situations illustrate how the implicature of "I think p" arises.

(5) One ordinary day Paul gets up in the morning and before he goes to work he looks out of the window and sees that it is raining. At that moment his roommate, unaware of the weather outside, asks if he should wear his coat today. Paul answers: "Yes you should, it's raining".

(6) One ordinary day David gets up in the morning and listens to the weather forecast on the radio, where they predict showers of rain in the afternoon. David takes his coat, expecting to be needing it later today, and goes to work. In the afternoon one

[11] Rosenthal 2005b

of David's colleagues calls it a day and David sees he wants to leave without a coat on, so he says: "You might want to put a coat on, I think it's raining."

Now Rosenthal would say that in the example in (5) Paul is verbally expressing his thought that it is raining and in the example in (6) David is reporting on his thought that it is raining, both of them getting across the message that it is raining. But the examples above clearly illustrate that this is not the case. Let us examine them individually.

The example in (5) is quite straightforward. Because Paul is looking out of the window and thereby seeing for himself that it is raining at the time someone asks him about the weather, he can say with certainty that it is raining. Paul knows that it is raining, which means that he can verbally express this knowledge properly by uttering "It's raining". The hearer of this utterance, in this case Paul's roommate, can interpret this utterance solely in one way: that it is raining. Of course, when concluding that that is the only interpretation we are assuming the co-operative principle is being observed. In the situation sketched in (5), when saying "It's raining" Paul is obeying the Maxims, and especially the Maxim of Quality: he believes what he says to be true and he has adequate evidence for it. If the hearer, his roommate, then assumes that Paul is being co-operative, he can only assume that Paul is being true and has adequate evidence for what he says, leading to him assuming that it is raining.

In the example in (6), on the other hand, we see a lack of knowledge about the weather at the time of the utterance. It does, however, concern some foreknowledge. When David sees his colleague leave without a coat and makes his comment about the weather he does not know the exact situation about the rain at that time, since he has not been able to look outside. So David does not know whether it is raining. But he does know, since he heard the weather forecast that morning, that there is a good chance that it is raining. This means that David believes (thinks) it is probably raining outside and he can properly verbally express this belief by uttering "I think it's raining". If we assume David is being co-operative he is obeying both the Maxims of Quantity and of Quality: he is making his contribution as informative as is required and he is not saying what he believes to be false or what he lacks adequate evidence for. Now the hearer, David's colleague, can, by assuming that David is being co-operative, infer that David does not know for a fact whether it is raining outside, but since he does mention his suspicion about the rain he must have reason to believe there is a good chance that it is raining outside.

Recall that Rosenthal would say that in the example in (5) Paul is verbally expressing his thought that it is raining and in the example in (6) David is reporting on his thought that it is raining, both of them getting across the message that it is raining. Following Grice and his co-operative principle, on the other hand, we would say that Paul is verbally expressing his knowledge that it is raining, getting across the message that it is raining, and David is verbally expressing his suspicion that it is raining, getting across the message that it might be raining. This contrast is shown clearly in the scheme below.

Utterance	*Rosenthal*	*Grice*
"It's raining"	expressing thought that it's raining	expressing knowledge that it's raining
"I think it's raining"	reporting thought that it's raining	expressing suspicion that it's raining

We can observe that as for the first utterance ("It's raining"), which is used in the example in (5), Rosenthal and Grice would both agree which message the speaker is intending to get across. But we cannot ignore that each of them reaches that conclusion in a significantly different manner. Rosenthal is not taking into account the actual circumstances and the conduct of conversation at all and is solely relying on thoughts and how we can express them. The only way he reaches that conclusion is by assuming that having a thought that it is raining and wanting to express that thought means that you want to get the message across that it is raining, whether it is actually raining or not. Grice does take into account the actual circumstances and the conduct of conversation which, assuming the co-operative principle, leads him to conclude that the speaker must know that it is raining when he is getting the message across that it is raining.

With the second utterance ("I think it's raining"), used in the example in (6), this different approach both philosophers are using results in them not even reaching the same conclusion anymore with respect to the message the speaker is intending to get across. Rosenthal still does not take into account the actual circumstances and the conduct of conversation and is just interpreting this utterance as a different way of expressing the thought that it is raining, which he calls reporting. According to him, the message is still that it is raining. Grice once again does take into account the actual circumstances and the conduct of conversation which, still assuming the co-operative principle, now leads him to conclude that the speaker does not know whether it is raining and just has a suspicion.

3.4 The implicature of "I think *p*"

Although Rosenthal explicitly defines the use of 'to think' in his HOT theory as having mental states, it seems that it just is not how we generally use that verb. Or at least not in utterances like the ones Rosenthal proposes us. But why is that? How does it come to be that in those kind of utterances our interpretation of the verb 'to think' is not the conventional interpretation?

To see where this comes from we can take a look at some different usages of 'to think' and see what makes the difference.

Utterance:	*Meaning of 'to think':*
"I am thinking about you"	to have a mental state
"I am thinking of a plan"	to have a mental state
"I can't think of his phone number"	to have a mental state
"I think so"	to have a belief/opinion
"I think (that) it's raining"	to have a belief/suspicion
"I think (that) that's a lovely sweater"	to have a belief/opinion
"I think (that) my husband's at work"	to have a belief/suspicion

When looking at the scheme above, we can discriminate two different ways of using 'to think': to think about/of something and to simply think something. Let us examine these two ways separately.

To think about p or to think of p is the way of using 'to think' that brings about the conventional interpretation of the verb in question. When you use 'to think' with an adverb like 'about' or 'of' you actually mean to say that you have a mental state with p as its content, where p can be anything you can think of (literally and figuratively). It does not matter whether p is true or false or even logically possible, since basically we are able to have any thought we want.

To think (that) p is the way of using 'to think' that implicates a different meaning for the verb than the conventional meaning of having a mental state. When you simply use 'to think p' the implicature made is that you have something like a belief, an opinion or a suspicion concerning p, where p is a proposition. This proposition can be either true or false (like "it's raining" or "my husband's at work") or something subjective (like "that's a lovely sweater"). This way, by predicating the proposition p with "I think" you can express the belief or opinion you have about p.

So the difference between the two ways of using 'to think' lies in the addition of an adverb like 'about' or 'of'. When those adverbs are present we interpret 'to think' as to actually think about something, but when they are not present we automatically assume the co-operative principle and interpret the utterance accordingly. That is also why it is so hard for us to interpret utterances like "I think it's raining" the way Rosenthal wants us to: as to have a mental state that it is raining. It is simply not the common use and interpretation of the verb 'to think' when used like that.

3.5 Performance conditions

Rosenthal speaks of 'performance conditions' for uttering a sentence, by which he means the conditions that define whether it is appropriate to utter the sentence. We must take notice that he does not explain exactly what these conditions are. He does, however, state that verbal expressions and reports of mental states are performance conditionally equivalent. By this he means that, for example, whenever it is appropriate for me to say "It's raining" (the verbal expression), it is also appropriate for me to say "I think it's raining" (the report). So, as far as Rosenthal is concerned, you can say "It's raining" or "I think it's raining" whenever you have a thought that it is raining.

But after discussing Rosenthal's inadequate way of interpreting 'to think' we can immedi-

ately see that this also causes problems for his statement about the supposedly performance-conditional equivalence between verbal expressions and reports. If we assume the co-operative principle you can only say "It's raining" whenever it actually rains and you know that it actually rains. If, on the other hand, you are not sure whether it rains and it is possible that it is actually raining, but it is also possible that it is not raining, you can say "I think it's raining". Obviously the performance conditions for uttering these two sentences are not the same, since when it would be appropriate to say "It's raining" it would not be appropriate to say "I think it's raining" and vice versa. You can also see this in the scheme below.

Utterance	*Performance conditions*
"It's raining"	It is actually raining You know that it is raining
"I think it's raining"	It could be raining It could not be raining You do not know whether it is raining

However, even though it seems here that verbal expressions and reports of mental states are not performance conditionally equivalent, we must not forget that we have just established that the report we are considering here is not really a report of the mental state in question: the utterance "I think it's raining" does not report a thought that it is raining, it expresses a belief in the possibility of rain. Since the report considered here is actually not a report of the mental state we cannot yet conclude that verbal expressions and reports of mental states are, unlike what Rosenthal claims, not performance conditionally equivalent.

The question now is if there are other ways to report a certain thought than just taking the wrongly used utterance "I think p" and if so, are they performance conditionally equivalent to the verbal expressions of the mental states they report? These other ways to report a mental state do exist. Here are some examples for reporting the thought that it is raining.

(7) "I'm having a thought that it's raining"

(8) "I have a mental state with 'it's raining' as its content"

Both examples in (7) and (8) are examples of proper reports of the thought that it is raining. It is the case for both sentences that when you utter them you are not saying anything about the truth value of or your knowledge or belief about a proposition that it is raining, you are solely and straightforwardly saying that you have a thought that it is raining. However, both utterances, and especially the utterance in (8), sound a little diffuse and unnatural. Indeed we would rarely use utterances like the ones in (7) and (8). But nevertheless they are proper reports of the thought that it is raining, which means we have something to work with.

Now that we have examples of proper reports of the thought that it is raining we can compare them with the verbal expression of the thought that it is raining and see if they have the same performance conditions. But first we have to make sure we are dealing with a proper verbal expression of the thought that it is raining as well. Rosenthal points out "It's raining" as a verbal expression of the thought that it is raining, yet we have established that by saying "It's raining" you are verbally expressing your *knowledge* that it is raining. This does, however, not exclude the possibility of it also still being a verbal expression of the thought that it is raining, since it would be very plausible to say that 'knowing that p' also entails 'thinking about/of

p': when you know something you also have a thought about it. Moreover, Rosenthal's own explanation for pointing out utterances like "It's raining" as verbal expressions of mental states is quite satisfactorily and hard to refute:

> If one says something meaningfully and sincerely, one thereby expresses some thought that one has.

With a proper verbal expression and a proper report at hand we can go on and compare their performance conditions to see if they are equivalent. When referring back to the aforementioned rare use of the utterances in (7) and (8) and the two-sided expression of "It's raining" it is relatively easy to finish the argument here. Rosenthal stated that verbal expressions and reports of mental states are performance conditionally equivalent and that that equivalence is second nature to us. But there is still no performance-conditional equivalence to be found, as shown in our new scheme below.

Utterance	*Performance conditions*
"It's raining"	It is actually raining You know that it is raining You have a thought about it
"I have a mental state with 'it's raining' as its content"	You have a thought that it is raining It could be raining It could not be raining

When verbally expressing your thought that it is raining you are not solely expressing your thought but, when observing the co-operative principle, also and moreover expressing your knowledge that it is raining. When reporting that thought, on the other hand, you are being more diffuse and thereby not really saying anything about whether it is raining or not or about any knowledge you may have. You only say that your thinking about it, which is exactly what a proper report should do.

So verbal expressions and reports of mental states are not performance conditionally equivalent. Moreover, since proper reports of mental states like the ones in (7) and (8) sound quite unnatural we can even say that if there would have been a performance-conditional equivalence between verbal expressions and reports of mental states it would definitely not feel like second nature to us: we hardly actually use those reports.

4 Consequences for the HOT theory

The concept of HOTs and how they arise in the HOT theory is something that Rosenthal completely explains through language and how we use it. However, we have seen that while relying so thoroughly on the use of language there is not much of what we have learned from pragmatics taken into account, which causes some major difficulties for the design of the HOT theory. The next sections will shed light on the consequences of these difficulties for the HOT theory of consciousness.

4.1 Explaining HOTs through language

We have noticed that Rosenthal puts a lot of emphasis on the distinction we have to make between verbal expressions and reports of mental states. He sees this distinction as essential to his HOT theory:

> There can be no doubt that saying that I think it's raining reports the thought that it's raining. If, in addition, that speech act also expressed that thought, the very distinction between expressing and reporting would collapse. We then could not infer from an intentional state's being conscious to the occurrence of a HOT.[12]

However, we have come to the conclusion that the kind of utterances Rosenthal points out as reports of thoughts, like "I think it's raining" reporting the thought that it is raining, actually are not reports of those thoughts. This is of course a major demerit for the theory considering that verbal expressions and reports are appointed as the fundaments of the theory. And as we have seen, this is not the only point where Rosenthal is being inaccurate when he concerns himself with the use of language. From the wrong use of the verb 'to think' and thereby producing improper reports he goes on to drawing wrong conclusions regarding subjects like interpretation and performance conditions.

The basic concept of the HOT theory, that what makes a mental state conscious is a HOT about that mental state, is entirely being explained and verified by means of the use of language: verbally expressing thoughts, reporting thoughts and from there on recognizing the occurrence of a HOT. But since we have established that we do not actually use language the way Rosenthal needs us to for his HOT theory to be explained and to work, we are not able to verify that basic concept of the HOT theory: that what makes a mental state conscious is a HOT about that mental state.

It should be obvious by now that the actual problem here is the attempt to explain the theory through the use of language in the first place. How we use language and convey thoughts, messages, knowledge, suspicion, doubt, etc. to one another is often multi-interpretable and subject to change. Therefore it is not a stable or suitable way to explain (let alone verify) a theory of consciousness like the HOT theory. Furthermore, the subject of the HOT theory, our thoughts, is something that goes on in the brain of each and every one of us and of which we all do not even speak the majority of the time. Millions of thoughts just stay in our head and never come out.

[12] Rosenthal 2005c

4.2 The concept of HOTs

Since we cannot be convinced of the HOT theory of consciousness by the way Rosenthal explains it and attempts to verify it we are left solely with the basic concept of the HOT theory. Of course, its plausibility is a bit affected and has become questionable, since the only explanation for the concept that we have turned out to be an incorrect explanation. This does, however, not mean that we cannot start over with just that basic concept and try to find out if it is a promising concept nevertheless.

Rosenthal stated that a mental state is conscious when it is accompanied by a suitable HOT about that mental state. What do we think of the concept that some thought you may have, say, that your neighbor's music is too loud, is a conscious thought when you have another thought about that thought? Is it a tenable concept? Is it in some other way than by means of the use of language explainable? Is it perhaps even verifiable in some way?

Whatever the answers to those questions may be, it has become obvious by now that they should not be sought in the field of linguistics like Rosenthal did. Fortunately, there are other options to explore. Since the subject of the HOT theory are thoughts, and thoughts arise and reside in our brains, it could be a logical option to try and find the answers in the field of neuroscience and psychology. Perhaps looking at brain activity during the occurrence of thoughts in a person's head and letting people analyze their own thoughts could help in finding out whether the concept of HOTs actually exists in our heads, right between our ordinary thoughts and everything else that goes around up there.

5 Conclusion

Within the field of the higher-order theories of consciousness the philosopher David Rosenthal has introduced the higher-order thought theory of consciousness. The basic concept of this theory is that what it is that makes a mental state conscious is a higher-order thought (HOT), and conscious mental states differ from those which are not in them being accompanied by a suitable HOT.

To explain how this concept works Rosenthal makes a distinction between verbally expressing a thought and reporting a thought. When you can report a thought it means that you have a HOT about that thought, since by reporting that thought you are implicitly verbally expressing the HOT about that thought. But since, according to Rosenthal, verbal expressions of thoughts and reports of thoughts are performance conditionally equivalent a thought is also accompanied by a HOT when you can simply verbally express it.

When explaining his theory by means of verbal expressions and reports of thoughts, with those two subjects being the skeleton of the theory, Rosenthal is considering the wrong kind of reports by using the verb 'to think' inaccurate and not in accordance with pragmatic theories like the co-operative principle. This leads him to being very persistent throughout the whole theory in not taking into account how we actually use language.

We see that the basic concept of the HOT theory, that what makes a mental state conscious is a HOT about that mental state, is entirely being explained and verified by means of the use of language. Unfortunately, we do not actually use language the way Rosenthal needs us to for the concept of HOTs to be explained. That the explanation of the theory by means of the use of language does not work is not a surprise, though, since a theory of consciousness should not be attempted to be explained by invoking language in the first place. How we use language is very multi-interpretable and subject to change, and therefore it is not suitable to explain or verify a theory of consciousness like the HOT theory.

Since the explanation of the basic concept of the HOT theory is far from satisfiable, the only thing that is left is that basic concept itself: that a mental state is conscious when you have a HOT about it. The answer to the question of whether this is a promising concept, however, should thus not be sought by revoking the use of language. The possibilities to find an answer could, however, lie in the field of neuroscience and psychology, since the subject of the HOT theory, our thoughts, is something that goes on in our brains. Moreover, most of the time our thoughts do not come out of that head of ours at all. Just imagine, can you even count how many thoughts you have kept to yourself over the last few days?

References

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts*. New York: Academic Press.

Moore, G. E. (1942). A reply to my critics. In P. A. Schilpp (Ed.), *The Philosophy of G. E. Moore*. LaSalle: Open Court.

Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan, and G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge : MIT Press.

Rosenthal, D. M. (2005a). Moore's paradox and consciousness. In *Consciousness and Mind*. New York : Oxford University Press.

Rosenthal, D. M. (2005b). Thinking that one thinks. In *Consciousness and Mind*. New York : Oxford University Press.

Rosenthal, D. M. (2005c). Why are verbally expressed thoughts conscious? In *Consciousness and Mind*. New York : Oxford University Press.