

UDK 165.81

13

1Rosenthal, D. M.

1Gennaro, R.

<https://doi.org/10.31337/oz.74.2.2>

Izvorni znanstveni rad

Primljeno: 23.8.2018.

Prihvaćeno 16.4.2019.

## Higher-Order Thought and the Problem of Shifting Subjectivity

Madhu Mangal Chaturvedi\*, A. V. Ravishankar Sarma\*\*

### *Abstract*

*In this paper, we argue that the two versions of the higher-order thought theory of consciousness, viz. Rosenthal's extrinsic higher-order thought and Gennaro's intrinsic higher-order thought, fail to explain the subjective character of a conscious mental state. Both these theories face what we call the problem of shifting subjectivity. Since these theories explain the consciousness of mental states in terms of a representational relation between two unconscious mental states with the help of a two-tiered representational structure divided into a higher-order thought and a lower-order mental state (which is the target of the higher-order thought), they fail to explain the subjective character of conscious mental states, which is intrinsic to them. In their account, the subjective character intrinsic to conscious mental states seems to shift from the target state to the higher-order mental state, which is separate from the target state. The objection is strong against Rosenthal's extrinsic higher-order thought theory, which clearly makes a distinction between a world-directed mental state and the higher-order thought representing it. However, although Gennaro's intrinsic higher-order thought theory is an attempt to preserve the intuition that consciousness is intrinsic to conscious mental states, it faces the problem of shifting subjectivity in the case of introspective consciousness.*

*Key words: consciousness, subjectivity, transitivity principle, HOT, EHOT, IHOT, introspection*

\* Madhu Mangal Chaturvedi, Ph.D., Assistant Professor, School of Philosophy & Culture, Faculty of Humanities and Social Sciences, Shri Mata Vaishno Devi University, Sub-Post Office Katra, J & K, India – 182320. E-mail: vijai.madhu@smvdu.ac.in

\*\* A. V. Ravishankar Sarma, Ph.D., Associate Professor, Department of Humanities and Social Sciences Indian Institute of Technology Kanpur, Kanpur, UP, India – 208016. E-mail: avrs@iitk.ac.in

## *Introduction*

Among the various representational theories of consciousness, Rocco J. Gennaro (1996; 2006; 2008; 2011) has put forth his claim that we can explain consciousness as an intrinsic property of conscious mental states, contrary to Rosenthal's higher-order thought (HOT) theory (1986; 1997; 2005), which takes consciousness to be an extrinsic feature of a conscious mental state. Rosenthal argues that by taking consciousness to be an extrinsic property of conscious mental states, his HOT theory provides a non-trivial, analysable account of conscious mental states. If consciousness is an extrinsic property, then it is also a relational property. A mental state is not intrinsically conscious. A mental state is primarily unconscious and becomes conscious by virtue of its intentional relation with a higher-order mental state directed at but distinct from it. For example, most of our beliefs are unconscious mental states which become conscious when they become the object of an appropriate higher-order mental state. According to Rosenthal, the appropriate higher-order mental state to represent the lower-order mental state must be thought-like. Since Rosenthal defends the view that the higher-order mental state must be a thought, and considers this higher-order thought to be extrinsic to a conscious mental state, his version of HOT theory can be called Extrinsic HOT (henceforth EHOT).

Gennaro, on the other hand, presents his Wide Intrinsicity View (WIV) theory as a better version of HOT than Rosenthal's in that it preserves the intuition that consciousness must be an intrinsic feature of conscious mental states. According to Gennaro's version of HOT theory, a conscious mental state is a complex mental state, and it is rendered conscious by virtue of its intentional relation with a Meta Psychological Thought (MET) directed at it but which is not distinct from the mental state. According to Gennaro, a conscious mental state is a complex of two constituent parts: 1) MET; and 2) a world-directed mental state. Thus, in contrast to Rosenthal's EHOT, Gennaro's version of HOT theory can be called Intrinsic HOT (henceforth IHOT).

Despite their differences on the structure of a conscious mental state, both Rosenthal and Gennaro agree that a higher-order representational state, directed at the target lower-order mental state, must be a thought and that it must be unconscious. In other words, a mental state is conscious by virtue of an intentional relation between two unconscious mental states. The problem with this approach is that it leaves the subjective character of conscious mental states, which is an inner awareness of being in a conscious mental state, unexplained. In other words, the problem for both versions of HOT is to explain how an intentional relation between two unconscious mental states (one of higher-order and the other of lower-order) results in rendering the lower-order mental state conscious.

Any good theory of consciousness must explain the subjective character of conscious mental states effectively and efficiently. Conscious mental states have a subjective character. Consciousness cannot be explained without the subjective

character of experience, and the subjective character of experience cannot be explained by a reductionist strategy. This subjective character of experience must be intrinsic to the conscious mental state. The best candidate for this subjective character is the inner-awareness of that conscious mental state. A mental state is conscious because it is subjective, and this subjectivity is the inner awareness that the subject has of being in that mental state. The subjective character is inbuilt into the intentional structure of consciousness. If conscious mental states are intentional in nature, then they have a correlational structure. This means that an intentional mental state is about something, i.e., it has an object, but it is also of someone, i.e., it has a subject. The subjective character of one's conscious mental state is the inner-awareness of being conscious of the object of one's mental state. For instance, when I am looking at the computer monitor there is an implicit awareness that I am conscious of the computer monitor in front of me. Similar to the objects of visual experience, we are aware of the objects of other sensory modal perceptual consciousness. This is the case of outer-world-directed mental states. Here my consciousness is implicitly directed at some object or the other in the external world. We say that consciousness is implicitly directed because we are not always self-aware when our consciousness is directed to objects in the external world. However, to be conscious of an object we must direct our consciousness to it. Now, similar to world-directed mental states we also have inner-directed mental states. When I am conscious of visually experiencing the computer monitor in front of me, I have my consciousness directed at the visual experience itself.

Rosenthal's EHOT faces the problem of shifting subjectivity in the case of world-directed conscious mental states, whereas Gennaro faces the problem of disappearing subjectivity by denying that world-directed conscious mental states are accompanied by an inner-awareness. The problem is that Gennaro's IHOT, albeit being certainly an improvement on Rosenthal's EHOT, does not successfully account for subjective character as an intrinsic feature of conscious mental states in the case of introspection.

Both EHOT and IHOT rely on the Transitivity Principle (TP). According to Rosenthal, TP means that »mental states are conscious only if one is in some way conscious of them« (Rosenthal, 2005, p. 4). TP is crucial to and the basis of the HOT theory of consciousness. The awareness of being in a conscious mental state is expressed in the form of a thought to the effect that "I am in a mental state" (Gennaro, 2006, p. 222). However, in both EHOT and IHOT this thought is unconscious in the case of world-directed mental states. In the case of introspective consciousness, the HOT is conscious, but only by virtue of being an object of another, unconscious HOT.

In Section (2) we discuss Rosenthal's extrinsic version of HOT theory. In Section (3), we discuss the problem of intrinsicity, which leads, in Section (4), to a discussion of Gennaro's intrinsic version of HOT theory being an improvement upon Rosenthal's EHOT. In Section (5), we present the problem of shifting su-

jectivity against EHOT and IHOT, and our concluding remarks appear in the Conclusion.

However, before moving on, a brief on the problem of subjectivity vis-à-vis reductionist–antireductionist debate on the nature of consciousness is in order. This will not only help in understanding the problem of shifting subjectivity raised against the two versions of higher-order thought theories of consciousness but also will highlight the reasons why philosophers turn to representationalism for a reductionist account of consciousness. To this we now turn.

### *1. The Problem of Subjectivity*

*Phenomenal Consciousness* presents a serious challenge to the world view endorsed by the natural sciences. If natural science aims at a comprehensive theory of reality, then it must also provide a suitable explanation of the phenomenon of consciousness. However, the major difficulty lies in the fact that phenomenal consciousness is not open to any objective methodology, as those employed in the natural sciences. One of the reasons is that, unlike any other physical phenomenon, phenomenal consciousness is subjective in nature. Despite various philosophical accounts of phenomenal consciousness, there is still something mysterious about this subjective character of consciousness which makes it difficult to analyze with the help of scientific methods, which basically involve observation and experiment. It is inevitable that both philosophers and scientists who are interested in understanding the phenomenon of phenomenal consciousness cannot ignore its essential feature: *subjectivity*.

Many contemporary philosophers argue that a merely physical description of reality may not be sufficient because it does not accommodate subjective phenomenal consciousness. The problem of subjectivity, therefore, is precisely this: How can phenomenal consciousness occur in the universe and yet remain irreducible to its fundamental physical constituents? In short, subjectivity is supposedly one last hurdle in providing a complete physicalist description of the world.

However, philosophers who argue for a representational theory of subjectivity are of the view that the natural explanation of subjectivity is possible. These representational theories are natural theories which may not reduce the subjectivity of phenomenal consciousness directly to physical terms. In the same vein, higher-order representational theories attempt a mentalistic reduction of consciousness. That is to say, they try to explain consciousness in terms of representational relation between two mental states and thus avoid the need to explain consciousness directly in physical terms. Once mentalistic reduction is complete, neural mechanism of higher-order representation too can be discovered to complete the reduction of consciousness in physical terms.

Now, there are two different views on the problem of subjectivity depending upon the definition of subjectivity accepted by their respective adherents. In the first group there are those who define subjectivity as qualitative character of a conscious mental state. For example, feeling of pain has distinct qualitative cha-

racter. In the second group, there are those who differentiate between the qualitative and subjective character of a conscious mental state. In the first group there are reductionists as well as antireductionists. For instance, antireductionist philosophers like Nagel (1974) and Chalmers (1996), claim that subjectivity as the qualitative character of experience cannot be reducible to the physical. According to Nagel and Chalmers, it is qualitative character which makes phenomenal consciousness irreducibly subjective.<sup>1</sup>

However, the first-order representationalists and reductionists like Dretske (1995) and Tye (1995), agree with the antireductionists that subjectivity is the qualitative character of experience, but argue for a claim that qualitative character is reducible in representational terms. First-order representationalists argue that all conscious mental states are representational and their qualitative character is the same as their corresponding representational content. The strategy adopted by representational theorists is to show that subjective qualitative character is not something in the head, but is identical with content which is external to the mind. Thus, these theorists make qualitative character a property of the object which exists in the external world. For example, the phenomenal experience of colour is the same as the surface reflectance property of the object being represented through visual perception.

Contrary to the first-order representational theorists, higher-order representationalists differentiate between qualitative character and the subjective, and define subjective character as *self-awareness*. On the one hand, first-order representationalists consider only qualitative character as a subjective feature, while on the other hand, higher-order representationalists provide a separate theoretical explanation for qualitative character and subjectivity. Karen Neander (1998) calls this strategy of higher-order representationalism the “division of phenomenal labor”. Rosenthal (1997), Lycan (1996), Carruthers (2000; 2005) and Gennaro (1996; 2011) are other advocates of higher-order theories of subjectivity, with variations in their respective higher-order representational theories. Uriah Kriegel (2005; 2009) has advanced the representational theories of subjectivity a step further to develop a self-representational account of the subjectivity of phenomenal consciousness. This theory also attempts to account for the qualitative character of the same within one single theory. The higher-order representationalists explain the subjectivity of mental states by an intentional relation between two numerically different mental states wherein a higher-order mental state is directed at the first-order mental state. According to the self-

1 Against this view, we can argue that subjectivity is not just merely to be treated as having qualitative character, because intentional states are also subjective in the sense that they belong to a conscious subject; or, in other words, the subject is aware of being in that state. For instance, Mary, the protagonist of Jackson’s knowledge argument, lacks qualia all the while when she is in her black and white room. However, she does have intentional states, such as beliefs and thoughts which she is aware of being in. Her beliefs and desires are still subjective states because she is self-aware of being in them.

representationalists, however, the intentional relation is inbuilt into the mental state itself.

Our target in this paper is higher-order thought version of representational theory of consciousness of which Rosenthal's EHOT and Gennaro's IHOT are the two versions. Since we argue that these two versions face the problem of shifting subjectivity and because Gennaro's IHOT is a modified version of Rosenthal's EHOT, we turn to his theory first.

## 2. *Rosenthal's EHOT Theory*

The definition of a conscious mental state, according to the EHOT theory of consciousness can be stated as follows: »a given mental state M is conscious if, and only if, the subject of that state possesses a thought about M, a thought to the effect that she has, or is in, M« (Rowlands, 2001, p. 290).

Philosophers (Carruthers, 2000; 2005; Gennaro, 1996; 2004; 2012; Rosenthal, 1986; 1991; 1997; 2005; 2009) present different versions of the HOT theory of consciousness. In this paper, we focus on D. M. Rosenthal's version of HOT theory. Rosenthal (1986; 1991; 1997; 2005) defends an EHOT theory of state consciousness. The EHOT theory, briefly, is the thesis that a mental state is conscious if it is the object of an appropriate HOT. A conscious mental state is a first-order representational state that is represented by a higher-order representational state in a suitable way. The higher-order representational state, according to Rosenthal, is a cognitive state, or a thought. Cognitive states, like thoughts, are different from perceptual states, like visual and auditory perception. A mental state is conscious when the subject has the thought that she is herself in that mental state. Paula Droege explains why HOT must have a self-referential element. She writes: »A HOT must include a self-referential element in order to identify the particular mental state that the thought is about« (Droege, 2003, p. 32). She further quotes Rosenthal, stating that in order to have self-referential thoughts, we need not require a very robust notion of the self. All that is needed, as Rosenthal puts it (1997, p. 741) is: »a concept that allows distinguishing between oneself and other things [...] it need not imply that the self has some special sort of unity, or is a center of consciousness, or is transparent to itself, or even that it has mental properties« (Droege, 2003, p. 33).

Not just any kind of HOT can make us aware of being in a mental state. To be conscious of being in a mental state, one must have an immediate, assertoric and occurrent thought about the state. The immediacy of the thought is required to account for the intuition that nothing seems to mediate between our awareness and its object. That the thought must be assertoric means that one should not doubt, believe or desire that one is oneself in the mental state. That the thought must be occurrent means that the thought must actually occur, and not be merely dispositional. The HOT must be "roughly contemporaneous" with the state that it is about, not a day before or a week after (Droege, 2003, p. 34). Rosenthal notes: »When a mental state is conscious, it is not simply that we are conscious of

the state; we are conscious of being in that state. This places a constraint on what the content of these HOTs must be; their content must be that one is, oneself, in that very mental state« (Rosenthal, 1997, p. 741).

Rosenthal (1986; 2003, pp. 326–327) observes that we are usually not conscious of being in a mental state, and his EHOT theory explains this fact. We are not conscious of being in a mental state because it is not accompanied by a HOT. Moreover, according to Rosenthal (1986; 1997), EHOT theory explains the difference between introspective and non-introspective consciousness. There are two kinds of conscious mental states: one in which the mental states are just conscious, and the other in which we are introspectively conscious. When we are non-introspectively conscious of a mental state, like in our everyday affairs, the HOT accompanying the mental state is unconscious. If someone is in a non-introspective conscious mental state, then he is not self-aware of being in that mental state. In the case of an introspective conscious mental state, one is aware of being in that mental state. Introspective conscious mental states are conscious mental states which are accompanied by conscious HOT. For introspection to occur there has to be a third-order thought which renders the second-order thought that accompanies the mental state conscious.

During introspection, one is aware of being aware of the mental state one is in. Introspection, according to Rosenthal, is the source of knowledge about our mental states, but only those of which we are conscious. We know our thoughts by introspective awareness. This immediate and non-observable awareness of our mental states is possible only if consciousness is intrinsic to them. But, according to Rosenthal, this view is indefensible. The view that consciousness is intrinsic to mentality is because we only know of conscious mental states. However, he argues, the way we know of certain things need not necessarily be a reliable source of the true nature of things.

Now, we discuss the problem of intrinsicality, which in essence explains the subjective character of conscious mental states which is intrinsic to them.

### *3. The Problem of Intrinsicality*

Since Rosenthal's higher-order thought theory explains consciousness in terms of an intentional relation between two distinct mental states, it suffers from the problem of intrinsicality. Goldman presents the problem of intrinsicality in the form of the following questions: »How could possession of a meta-state confer subjectivity or feeling on a lower-level state that did not otherwise possess it? Why would being an intentional object or referent of a meta-state confer consciousness on a first-order state? A rock does not become conscious when someone has a belief about it. Why should a first-order psychological state become conscious simply by having a belief about it?« (Goldman, 1993, p. 366).

»Our ordinary understanding of awareness or consciousness seems to reside in features that conscious states have in themselves, not in relations they bear to

other states« (Goldman, 1993, p. 367). Goldman calls this thesis *intrinsicism*. Intrinsicism about conscious mental states can take many forms. Intrinsicism does not necessarily mean that consciousness *prima facie* seems to be an intrinsic property of mental states. By intrinsic property, we mean that property of a thing which does not depend on its relation to something else. For instance, being a brother is not an intrinsic property, for it depends on a certain relationship between two individuals. In other words, an intrinsic property is not a relational property. This version of intrinsicism certainly makes the concept of consciousness unanalysable and circular. However, in another version, intrinsicism means that conscious mental states have some intrinsic property which distinguishes them from unconscious mental states. Traditionally, this intrinsic property has been identified as the feel or qualia attached to conscious mental states. However, primarily, qualia are associated only with a certain kind of conscious mental state, viz. phenomenal states. If we take phenomenal mental states as the paradigm of conscious mental states, we can still distinguish two features which define a phenomenal mental state. As highlighted by Uriah Kriegel (2005; 2009), phenomenal states have two distinguished features: qualitative character and subjective character. Regarding qualitative character there has been a lot of debate, but most philosophers would agree that qualitative character is not an intrinsic property of a phenomenal state but extrinsic to it. A phenomenal mental state has qualitative character by virtue of representing the phenomenal property of its object. In other words, the phenomenal character of a conscious mental state is individuated by the property of the object external to it.

However, subjective character gives a phenomenal mental state its “me-ness”. It is subjective character which is intrinsic to a phenomenal mental state. Following Kriegel (2005; 2009), we can define subjective character as the inner-awareness that an agent has of being in a phenomenal state. This inner-awareness, it could be argued, is something common to not just conscious phenomenal states, but to any kind of conscious mental state.

Now, consciousness, according to Rosenthal, is a relational property of mental states. This means that, according to HOT theory, a mental state becomes conscious by virtue of being an object of another numerically distinct mental state. Nevertheless, this view has intuitive difficulties. If a subject S is conscious of a mental state M because he has another mental state M1 about M, then he is conscious, rather, of M1 and not of M. Moreover, when I see a table, I am conscious of the table. Here, consciousness does not become a property of the table. In other words, when I am conscious of something, i.e., transitively conscious, the object of my consciousness is not itself conscious. This is the well-known difficulty for HOT theories and is called the “problem of the rock,” as mentioned by Gennaro (2006, p. 225). It is originally attributed by him to Goldman (1993) as quoted above.

In the above-mentioned case, there is a distinction between the object of consciousness, namely, the table, and the state of consciousness, namely, seeing. However, when I am conscious of being in a mental state, for example, seeing



the table by virtue of having a HOT about it, I am not actually in that state unless both the first-order state, i.e., the seeing, and the higher-order state, i.e., the thinking, occur simultaneously and are identical. However, since Rosenthal emphasizes on the distinctness of the higher-order representation of a mental state from the mental state itself for the sake of a non-circular definition of consciousness, it becomes difficult to admit that one is aware of being in a mental state by virtue of being in another mental state that represents it.

Awareness in general seems to be more primary than thought. We are aware of having thoughts. For example, Rosenthal writes, »For one to be aware of one's first-order states, these HOTs must have the content that one is, oneself, in the state in question« (Rosenthal, 2009, p. 165). This means that one has direct, immediate and non-inferential awareness of having the thought that one is in a certain mental state. This rather implies that one is actually aware of having a thought but not being in a state towards which the thought is directed. For example, Güven Güzeldere raises the following question against the higher-order perception model of consciousness: »What exactly is being perceived in the 'perception of what passes in one's own mind': the content of the mental state that happens to be 'passing through' one's mind at the time or the mental state itself? Or the content of another thought to the effect that one is having such a mental state?« (Güzeldere, 1997, p. 792).

Similarly, one can ask what exactly one is conscious of while having a thought about a mental state: the content of the thought or the mental state itself. Since Rosenthal emphasizes on the occurrence of an unconscious HOT directed at a target mental state in order to avoid infinite regress, we have no way to be conscious of the content of the thought unless there is another third-order thought directed at the second-order thought. Nevertheless, this happens only in the case of introspective awareness. The problem is: How can one explain awareness of being in a conscious mental state by positing an unconscious thought? One answer comes from Gennaro (1996; 2011), who proposes a slightly modified version of HOT theory, namely, the Wide Intrinsicity View, or Intrinsic HOT (IHOT) theory, which we discuss now.

#### 4. *Gennaro's IHOT*

Rocco J. Gennaro (1996) criticizes Rosenthal's view that consciousness is an extrinsic or relational property of mental states. Rosenthal's argument for a distinct HOT depends on showing that consciousness is an extrinsic property of mental states. Rosenthal argues that if we accept that consciousness is an intrinsic property of mental states, then we cannot provide a non-circular definition of mental state consciousness. This means that conscious mental states become simple and unanalyzable states.

Gennaro argues that consciousness seems to be an intrinsic property of mental states. A property of X is *intrinsic* if it does not depend on X being in relation to something else. For instance, "being a brother of someone" is not an intrinsic

property. Gennaro observes that being in a conscious mental state does not seem like an extrinsic property. So, he argues, a theory which explains this intuitive fact is preferable to a theory which denies it (Gennaro, 1996, p. 27). Moreover, according to Gennaro, »even if consciousness is an intrinsic property of some mental states, it surely does not follow that those states are simple or unanalyzable« (Gennaro, 2012, p. 57).

Unlike Rosenthal, instead of assuming a numerically distinct mental state, viz., an HOT, to explain the consciousness of a mental state, Gennaro proposes that a meta-psychological thought is built into the mental state itself. Gennaro terms his thesis as the “Wide Intrinsicity View” (WIV), as opposed to the “Narrow Intrinsicity View” (NIV), which he attributes to Nagel. NIV states that consciousness is a simple property of mental states which is non-relational, ineffable and intrinsic (Gennaro, 1996, p. 15). Since NIV stresses that our knowledge of conscious mental states is available from the first-person *point of view*, this means that conscious mental states are theoretically not analysable, and this is explanatorily unhelpful.

Gennaro’s WIV implies that a conscious mental state is a first-order representational state which is a world-directed mental state but which has consciousness intrinsic to it. So, Gennaro’s WIV treats a conscious mental state as a complex state in which the higher-order state is not distinct but forms a part of it. There is no higher-order state or meta-psychological state which represents the first-order state. Instead, the meta-psychological state is a constituent of the first-order mental state as a distinct part which it renders conscious. Effectively, thus, Gennaro’s MET is an intrinsic version of Rosenthal’s HOT (Gennaro, 2011, p. 55). This is why Gennaro has called his WIV an Intrinsic HOT in contrast to Rosenthal’s Extrinsic HOT.

»According to the WIV, what makes mental states conscious is *intrinsic* to conscious states, but a kind of *inner* self-referential and relational element is also present *within* the structure of such states. In contrast to standard HOT theory, the WIV says that *first-order* conscious mental states are complex states containing both a world-directed mental state—part M and an unconscious meta-psychological thought (MET)« (Gennaro, 2011, p. 55).

Gennaro, however, to an extent agrees with Rosenthal in asserting that the meta-psychological state is always a thought. Having the thought that one is in a mental state makes that state conscious, or rather, the fact that one is in a conscious mental state is identical with having a contemporaneous, non-inferentially caused thought that one is in that state. Nonetheless, for Gennaro, the thought, which is a meta-psychological state, is not a distinct state. Gennaro expresses his WIV thus: »A mental state M of a subject S is conscious if and only if S has a suitable (unconscious) meta-psychological thought M\* (= MET), directed at M, such that both M and M\* (= MET) are *proper parts* of a complex, conscious mental state CMS« (Gennaro, 2006, p. 222).

As far as the subjectivity of a conscious mental state is concerned, *prima facie*, Gennaro’s IHOT (WIV) account seems preferable to Rosenthal’s HOT theory.

Since Rosenthal's account admits the numerical distinction between the CMS and the HOT, it fails to explain the intimate relation between the subjectivity of a conscious mental state, i.e., the self-awareness of being in that state, and the state itself. Gennaro, on the other hand, admits no numerical distinction between the CMS and the HOT. Moreover, he asserts that »self-consciousness consists in having meta-psychological thought,« and also that »thinking about one's own mental states is definitive of self-consciousness« (Gennaro, 1996, p. 16).

So, for Gennaro, being in a conscious mental state is to have an unconscious meta-psychological thought which is directed at the mental state and is a part of the CMS. Here, the intimate relation between consciousness and subjectivity is explained in a better way as compared to the HOT theory. Firstly, the mental state and the MET are not distinct states but form part of a single complex conscious mental state. Secondly, being in a conscious mental state by this account entails being self-conscious by virtue of having a MET.

Because of the first feature, we can say that IHOT avoids "the problem of the rock".

This problem, as mentioned above, is the problem of explaining how simply by being represented by consciousness renders something conscious. In other words, being an object of consciousness is to be represented by it and when objects like rock are represented by consciousness they don't become conscious. Now in IHOT model, since the HOT is a part of the CMS, hence it renders the target state conscious. Furthermore, to be conscious of a mental state is not to be in a separate mental state, namely, a HOT. This account also avoids any consciousness-content distinction. The second feature explains the first-person perspective, or the self-awareness of being in a mental state. There is something it is like for the subject to be in a conscious mental state because it has a MET directed at it. This is precisely what subjectivity is.

However, the problem with this view is the relation between the MET and the mental state as the two parts of a single complex state. How are we to understand the relationship between the MET and the mental state at which it is directed? Gennaro views the relation between the MET and the mental state as the real relation, as opposed to being merely conceptual, a distinction introduced by Kriegel (2005). The relation between two parts of a single mental state is real if both parts are »essentially connected, or bound, in a certain way; there is a psychologically real relation or integration between the parts« (Gennaro, 2006, p. 239). On the other hand, the relationship is conceptual if we decide to treat them as related, i.e., either both of the parts of the CMS, the MET and the mental state, are merely present together but are not related with each other in some essential way, or both are integrated into one single mental state, eventually producing the CMS.

The question that is left open in both the approaches is the question concerning the ontological status of the CMS, on the one hand, and the MET and the mental state as parts of the CMS, on the other. The CMS comes into existence only after the MET and the target mental state are integrated. Hence, once the

CMS is produced, what happens to its parts? The mental state at which the MET is directed must be essential to the CMS because that is what becomes conscious.

### *5. The Problem of Shifting Subjectivity*

Both Rosenthal's Extrinsic HOT and Gennaro's WIV face problems when it comes to explaining the subjectivity of conscious mental states. Subjectivity is an essential feature of any conscious mental state. In the contemporary literature on consciousness, most philosophers use subjectivity and qualitative character synonymously and interchangeably.<sup>2</sup> However, as Levine assumes that qualitative character and the subjectivity of experience are two different aspects of the problem, the mystery shrouding consciousness is more due to its subjectivity and deeper than its qualitative character. The problem for the materialists, according to Levine, is to explain the qualitative character of conscious experience as well as the awareness of it (Levine, 2001, p. 7).

Uriah Kriegel contends that subjectivity is a kind of inner awareness which is a distinct feature of a mental state from its qualitative character (Kriegel, 2005). According to Kriegel (2005; 2009), phenomenal consciousness has the following two elements: inner awareness and qualitative character (Gulick, 2011). The qualitative character of conscious experience makes it what it is. As an example, we may consider the conscious experience of throbbing pain as different from that of burning pain. The difference in both the experiences of pain is in their qualitative character. However, Kriegel argues that pain can only be conscious if it has subjective character, i.e., the subject has an inner awareness of the pain experience. When I am looking at a red rose, there is something it is like for me to see a red rose and, at the same time, I am aware that I perceive a red rose. For Kriegel, this for-me-ness, or my awareness of the red rose, is the subjective character of the conscious experience that I have.

Now, this for-me-ness or inner awareness of being in a conscious mental state cannot be distinct from the conscious mental state. However, within the framework of HOT theories, the location of subjective character seems to be shifting because of their two-tiered structure of conscious mental states. Roughly, a mental state becomes conscious by virtue of being represented by a higher-order mental state. This representation takes the form of a thought to the effect that "I am in a mental state" (Gennaro, 2006, p. 222). In the HOT model of consciousness, this "I am in a mental state" thought is the equivalent of the inner awareness of being in a mental state. This implies that the subjective character

2 This is the view common to most of the philosophers, reductionists and anti-reductionists alike, as well as eliminativists, about phenomenal consciousness. See Chalmers (1996), Jackson (1986) and Nagel (1974; 1986), who are anti-reductionists. Among the reductionists are Dretske (1995) and Tye (1995). And the famous eliminativists are Churchland (1994) and Dennett (1991). However, Kriegel (2005; 2009) and Levine (2001) do make a distinction between qualitative character and subjective character of phenomenal consciousness.

of a conscious mental state, according to the HOT theory, is the thought that “I am in a mental state.”

The problem is that, given the two-tiered structure of the HOT model of consciousness, this thought is either unconscious when accompanying a lower-order state or is invariably located in the higher-order state rather than in the lower-order state which it is supposed to render conscious. This means that the subjectivity is shifting from the lower-order mental state to the higher-order mental state, hence the name of the problem of shifting subjectivity.

In fact, HOT theory does not really get to explain this crucial feature of conscious mental states.

The problem of shifting subjectivity is particularly damaging to Rosenthal’s EHOT, since it is strong even in the case of a world-directed conscious mental state. For IHOT, on the other hand, the objection is particularly strong in the case of introspective consciousness. In the case of introspective consciousness, Rosenthal’s EHOT theory posits that an unconscious HOT, which is about a world-directed mental state, is rendered conscious by another third-order HOT. In this scheme, we have three distinct mental states, first the world-directed mental state of which we become introspectively conscious, second, the unconscious HOT which renders it conscious, and third, another HOT which renders the unconscious HOT conscious. Now, EHOT explains how a world-directed mental state becomes conscious and how the same mental state is introspectively conscious. Rosenthal’s version, as we have seen, has a problem in the first-step itself due to the positing of a HOT distinct from the world-directed mental state.

Gennaro tries to improve on Rosenthal’s EHOT by introducing the MET as a part of the complex conscious world-directed mental state. However, this strategy fails in the case of introspective awareness. Since we are introspectively conscious of being in the same mental state, the MET should be a part of it. The intrinsically subjective nature of a conscious mental state, whether introspective or non-introspective, demands that the source of awareness of being in that mental state cannot be distinct from the mental state itself. However, in the case of introspection, Gennaro places the conscious MET and the unconscious MET rendering it conscious as parts of a complex mental state distinct from the mental state of which one becomes introspectively conscious (See Figure 1). This means that, according to the transitivity principle, the thought with the content that ‘one is introspectively aware of being in the mental state’ is distinct from the thought itself. In other words, the third-order thought that makes introspective awareness possible by representing the second-order thought is distinct from it. That is to say, introspective awareness and the higher-order thought making it conscious are distinct. This makes Gennaro’s IHOT vulnerable to the problem of the rock.

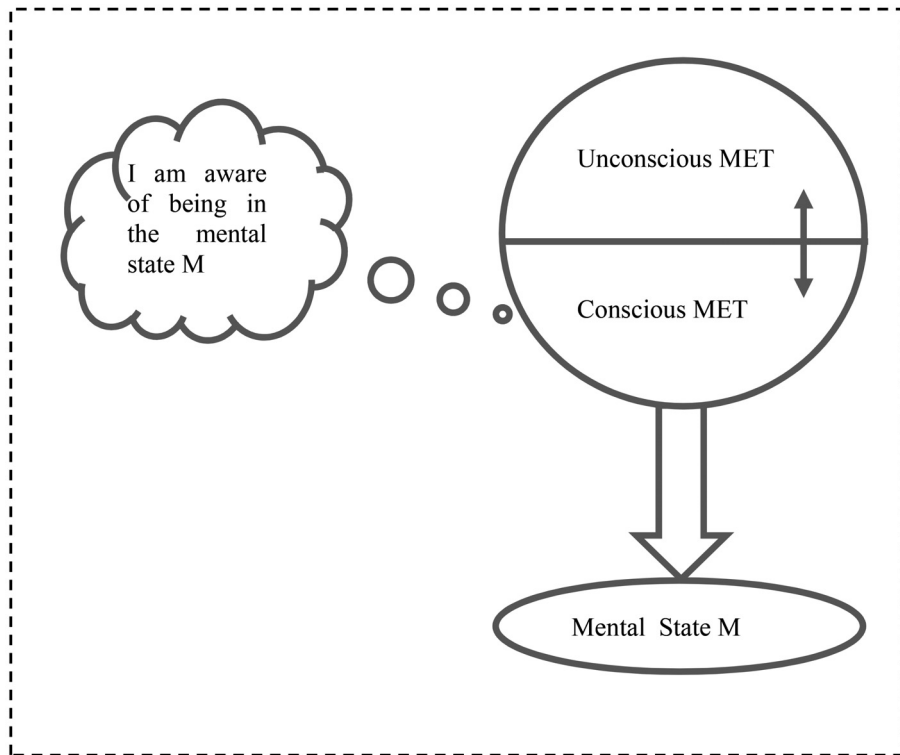


Figure 1. *Gennaro's IHOT Model of Introspection*

Nevertheless, Gennaro's IHOT avoids the problem of the rock in the case of world-directed mental states by making the MET and the world-directed mental state part of the one single conscious mental state. In this case, the conscious mental state is not an atomic state but includes a complex structure. However, Gennaro's IHOT account for world-directed mental states faces a different problem. The problem is to explain why an unconscious MET makes a mental state conscious by virtue of being a part of the same state. The major argument for allowing the MET, albeit unconscious, as an intrinsic but separate part of the complex mental state is to keep the theory of consciousness informative and render conscious mental states analyzable. Gennaro's IHOT does not provide a similar integrated model for introspective consciousness. Here, IHOT faces the problem of the rock.

Gennaro's response to the "problem of the rock" by using Lycan's strategy and pointing out those rocks, like any other non-psychological, physical objects in this world that we are transitively conscious of, cannot become intransitively conscious like a mental state which is the target of a HOT (Gennaro, 2012, p. 73). The idea is that mental states are ontologically different from any non-mental objects, like rocks, tables or chairs. Moreover, a rock does not become intransi-

tively conscious like a mental state. That is to say, we do not attribute creature consciousness to rocks or tables. We never say a rock is awake or a table is asleep (Carruthers, 2016). However, Gennaro interprets the problem of the rock as an objection, which states that higher-order representationalism proves too much.

»Clearly this objection could be devastating. If any HO theory succeeds, then it would be guilty of proving too much. That is, it would make too many things conscious and would thereby be reduced to absurdity. Thus a HO theorist must block the generalization to rocks and the like without sacrificing an informative analysis of consciousness« (Gennaro, 2012, p. 71).

Gennaro tries to diffuse the problem of the rock objection by pointing out that the case of rocks is not like that of mental states. Any non-mental entity, like a rock, does not become conscious by virtue of being represented by a world-directed mental state, but just becomes the object of a mental state. However, the intuitive appeal behind the problem of the rock is not just that, if right, it would show that HO theories prove too much. The intuitive appeal behind this objection is that even if consciousness is a property of psychological entities, like various kinds of mental states, then how do we explain the fact that a mental state becomes conscious by virtue of being an object of a meta-state? In other words, taking the analogy of the rock, when the rock is represented it does not become conscious, but it becomes the object of a mental state. In this intentional relation between a mental state and the rock, the subject pole is the mental state and the object pole is the rock. At no point in this relational structure does the rock become the subject, or in other words, the locus of consciousness.

Gennaro's strategy against the problem of the rock does not work in the case of introspection, since the problem with Gennaro's IHOT model of introspective consciousness is that it does not really explain the subjectivity of introspectively conscious mental states. Gennaro's IHOT solution for introspective consciousness, as against Rosenthal's EHOT, consists in making the introspected mental state (target state) an object of a complex conscious mental state distinct from the target state, which unlike Rosenthal's model has unconscious MET and Conscious MET as its two parts. The conscious mental state is a complex state having two parts within it, viz. the unconscious MET directed at the conscious MET. The idea is that during introspection the agent is in a distinct higher-order mental state which is directed at the mental state under introspection. Since introspective consciousness results from a representational relation between two mental states of different orders, the agent's awareness of the mental state being introspected by her is located in the higher-order mental state, the conscious MET which is conscious by virtue of being an object of another, albeit, unconscious MET. To maintain that consciousness is intrinsic to the mental state even in the case of introspection, the conscious MET and unconscious MET directed at it form part of a single complex mental state.

Rosenthal's EHOT and Gennaro's IHOT take a perceptual model of introspection (PMI). PMI takes introspection as structurally similar to external perception. Hence, the HOT and WIV models treat a mental state under introspection

as a lower-order mental state distinct from and represented by a higher-order mental state. This model suggests that introspective awareness is not located in the mental state being introspected. Moreover, this introspective awareness is also a result of the representational relation between an unconscious MET and a conscious MET which are combined and directed together at the target mental state. However, even during introspection the agent's introspective awareness is not distinct from the mental state she is introspecting. Phenomenologically speaking, the act of introspection and the object of introspection, when the object of introspection is a mental state, must be intimately related. Moreover, one can introspect a mental state of which one is aware of being in. This suggests that to introspect a mental state is more about attending to it while being aware of being in it, at least partially, rather than attending to it like a distinct mental state which one is not aware of being in. For being in an introspective state is also to be able to feel its contents, and one cannot feel contents of a mental state without being aware of being in it. For example, when we introspect causes of fear, we do feel what it is like to fear, without which introspection would not be possible.

The problem of shifting subjectivity is a serious challenge to HOT theories and goes deeper compared to other objections raised against it. We can demonstrate the significance of this objection with the help of the following two examples.

*Cognitive overload:* Cognitive overload or information overload problem raised by Carruthers (2000; 2005) is that we have a large number of first-order representations of our bodily and extra-bodily environment at any given moment. We watch and listen at the same time but not all of these representations are conscious. According to Carruthers, HOP and HOT theories assert that higher-order representations of the first-order representations render them conscious. However, this higher-order representational state, whether a thought or inner perception-like, should actually occur to target the first-order representations. Since first-order representations are responsible for the richness and detail of our experience, so the higher-order states must also have the richness of detail to explain the richness of the consciousness of the perceptual experience. This would strain our cognitive system, for it would require devoting more cognitive space and computational resources for re-processing the same information at two different levels, thus amounting to an information overload (Carruthers, 2005, p. 54).

*The problem of infinite regress:* The problem is that since consciousness, according to HOR theories, is a relational or extrinsic property of mental states, therefore, it is right to assume that an unconscious mental state becomes conscious by virtue of being the intentional object of another mental state. However, if an unconscious mental state M becomes conscious by virtue of being an intentional object of a conscious mental state M1, then we need to posit yet another conscious mental state M2 in order to explain the consciousness of M1, ad infinitum (Kriegel, 2005).

The cognitive overload and infinite regress problem are a result of failure of HOT theories to integrate subjectivity with the conscious mental state. Since Ro-



rosenthal and Gennaro explain consciousness as the result of the relation between the two mental states, they create a gap between the conscious mental state and its subjectivity at various levels. Rosenthal leaves the gap in the case of world-directed perceptual states as well as at the level of introspective consciousness, and for Gennaro it is clear at the level of introspective consciousness. In addition, the richness of first-order mental states to which Carruthers refers is a problem for HOT theories precisely because they posit separate higher-order mental states to account for the subjectivity of consciousness.

Moreover, the problem of shifting subjectivity goes deeper than the problem of intrinsicity or the problem of the rock raised by Goldman (1993), and in a way the latter is an extension of the former.<sup>3</sup>

### *Conclusion*

In the present paper, we have argued that the two versions of the HOT theory of consciousness, viz. Rosenthal's EHOT and Gennaro's IHOT, fail to explain subjectivity, a crucial feature of conscious mental states. Both versions face what we have termed as the problem of shifting subjectivity at various levels in explaining consciousness. Whereas Rosenthal's EHOT faces the problem of shifting subjectivity in explaining outer world-directed conscious mental states, Gennaro's IHOT encounters it in explaining introspective consciousness. The argument is that EHOT fails to explain subjective character as a feature of lower order states in their relational model of consciousness. It is the problem of shifting subjectivity which is an argument for the relational account of subjective character. In other words, subjectivity is an intrinsic feature of a conscious mental state, and if we try to explain it in relational terms, like the HOT theories do, we face the problem of shifting subjectivity. The location of subjectivity seems to shift from a lower-order mental state to a HOT, and so on.

### *Bibliography:*

- Carruthers, Peter (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Carruthers, Peter (2005). *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Clarendon Press.
- Carruthers, Peter (2016). Higher-Order Theories of Consciousness. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/>.
- Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

3 For detailed and easy discussion on the higher-order theories of consciousness and representational theories of consciousness in general please see Carruthers (2016) and Kriegel (2007). For an easy introduction to the problem of consciousness and reductionist-antireductionist debate see Chalmers (1996) and Tye (1995). Moreover, for a comprehensive discussion on theories of consciousness see Seager (2016) and Weisberg (2014).

- Churchland, Paul M. (1994). *The Engine of Reason, the Seat of the Soul: Philosophical Journey into the Brain*. Cambridge, MA: The MIT Press.
- Dennett, Daniel C. (1991). *Consciousness Explained*. Tavistock: Penguin Books.
- Dretske, Fred (1995). *Naturalizing the Mind*. Cambridge, Mass.: The MIT Press.
- Droege, Paula (2003). *Caging the Beast: A theory of sensory consciousness*. Amsterdam — Philadelphia: John Benjamins Publishing Company.
- Gennaro, Rocco (1996). *Consciousness and Self-Consciousness: A Defense of HOT Theory of Consciousness*. Amsterdam: John Benjamins.
- Gennaro, Rocco (2006). Between pure self-referentialism and the (extrinsic) HOT theory of consciousness. In U. Kriegel and K. Williford (Eds.), *Self-Representational Approaches to Consciousness* (pp. 221–248). Cambridge, MA: MIT Press.
- Gennaro, Rocco (2012). *The Consciousness Paradox: Consciousness, Concepts, and HOTs*. Cambridge, Mass.: A Bradford Book — The MIT Press.
- Goldman, Alvin (1993). Consciousness, Folk Psychology, and Cognitive Science. *Consciousness and Cognition*, 2, pp. 264–282.
- Gulick, Robert van (2011). Subjective Consciousness and Self-Representation. *Philosophical Studies*, 159(3), pp. 457–465.
- Güzeldere, Güven (1997). Is Consciousness the Perception of What Passes in One's Own Mind? In N. Block, O. Flanagan and G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 789–806). Cambridge, MA: MIT Press, 789–806.
- Jackson, Frank (1986). What Mary didn't Know. *Journal of Philosophy*, 83, pp. 291–295.
- Kriegel, Uriah (2005). Naturalizing Subjective Character. *Philosophy and Phenomenological Research*, 71, pp. 23–56.
- Kriegel, Uriah (2007). Philosophical Theories of Consciousness: Contemporary Western Perspectives. In P. Zelazo, M. Moscovitch, and E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 35–36). Cambridge: Cambridge University Press.
- Kriegel, Uriah (2009). *Subjective Consciousness*. New York: Oxford University Press.
- Levine, Joseph (2001). *The Purple Haze*. New York: Oxford University Press.
- Lycan, William G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Nagel, Thomas (1974). What Is It Like to Be a Bat? *Philosophical Review*, 83, 453–450.
- Nagel, Thomas (1986). *The View from Nowhere*. London: Oxford University Press.
- Neander, Karen (1998). The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness. *Philosophical Perspectives*, 12, Language, Mind, and Ontology, pp. 411–434.
- Rosenthal, David M. (1986). Two Concepts of Consciousness. *Philosophical Studies*, 49(3), pp. 329–359.
- Rosenthal, David M. (1991). The Independence of Consciousness and Sensory Quality. *Philosophical Issues*, 1, pp. 15–36.
- Rosenthal, David M. (1997). A Theory of Consciousness. In N. Block, O. Flanagan and G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 729–753). Cambridge, MA: MIT Press.
- Rosenthal, David M. (2005). *Consciousness and Mind*. Oxford: Clarendon Press.
- Rowlands, Mark (2001). Consciousness and HOTs. *Mind & Language*, 16(3), pp. 290–310.
- Seager, William (2016). *Theories of Consciousness: An Introduction and Assessment*. New York: Routledge.
- Tye, Michael (1995). *Ten Problems of Consciousness*. Cambridge, Mass.: The MIT Press.
- Weisberg, Josh (2014). *Consciousness*. Malden, MA: Polity.

*Misao višega reda i problem subjektivnosti koja se pomiče*

Madhu Mangal Chaturvedi\*, A. V. Ravishankar Sarma\*\*

*Sažetak*

*U radu autori iznose tvrdnju da dvije verzije teorije svjesnosti tzv. misli višega reda, Rosenthalova ekstrinzična misao višega reda i Genarrova intrinzična misao višega reda, ne obrazlažu subjektivni karakter svjesnoga mentalnoga stanja. Obje teorije susreću se s takozvanim problemom subjektivnosti koja se pomiče. S obzirom da te teorije tumače svjesnost mentalnih stanja pomoću dvoslojne reprezentacijske strukture podijeljene na misao višega reda i mentalno stanje nižega reda (koje je meta misli višeg reda), one ne uspijevaju obrazložiti subjektivni karakter svjesnoga mentalnoga stanja koji im je intrinzičan. Prema njihovu tumačenju, subjektivni karakter intrinzičan svjesnim mentalnim stanjima pomiče se, čini se, sa stanja mete na mentalno stanje višega reda, koje je odvojeno od stanja mete. Autori upućuju ozbiljnu primjedbu na Rosenthalovu teoriju ekstrinzične misli višega reda, koja jasno razlikuje svijetu usmjereno mentalno stanje od misli višega reda, koja ga predstavlja. Međutim, iako Gennarova teorija intrinzične misli višega reda predstavlja pokušaj očuvanja intuicije da je svjesnost intrinzična svjesnim mentalnim stanjima, ona se susreće s problemom subjektivnosti koja se pomiče u slučaju introspektivne svjesnosti.*

*Ključne riječi: svjesnost, subjektivnost, načelo tranzitivnosti, HOT, EHOT, IHOT, introspekcija*

\* Dr. sc. Madhu Mangal Chaturvedi, School of Philosophy & Culture, Faculty of Humanities and Social Sciences, Shri Mata Vaishno Devi University, Sub-Post Office Katra, J & K, Indija — 182320. E-adresa: vijai.madhu@smvdu.ac.in

\*\* Izv. prof. dr. sc. A. V. Ravishankar Sarma, Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur, Kanpur, UP, Indija — 208016. E-adresa: avrs@iitk.ac.in