Cognitive Phenomenology, Semantic Qualia and Luminous Knowledge

by

Neil Tennant*

Department of Philosophy
The Ohio State University

February 2, 2006

Abstract

This paper employs an account of game-theoretical semantics in a thought-experiment designed to make vivid and plausible the claim that there is something that it is like to understand a sentence as expressing a thought. Knowing that one grasps a sentence as meaningful is offered as a paradigm case of 'luminous' knowledge, in Williamson’s sense.

*To appear in Patrick Greenough and Duncan Pritchard, eds., Williamson on Knowledge, Oxford University Press, 2006. I would like to thank Adam Podlaskowski, Galen Strawson and Michael Tye for helpful correspondence.
1 Introduction

For Timothy Williamson [8], at p. 93, a cognitive home is a region of logical space in which ‘everything lies open to our view’:

... in our cognitive home...mistakes are always rectifiable... we are not omniscient about our cognitive home. We may not know the answer to a question simply because the question has never occurred to us. Even if something is open to view, we may not have glanced in that direction. ... the point is that such ignorance is always removable. (p. 94)

The anti-realist, for whom all truths are knowable, would here contemplate, for each knower, the rosy prospect of a spacious and well-furnished home, with creature comforts reflecting the tastes, interests and sensibilities of its owner. But, it turns out, the bourgeoisie are to be dispossessed. Epistemically, we are all street persons. It is not clear whether Williamson laments our condition, or points it out with grim satisfaction:

... we are cognitively homeless. Although much is in fact accessible to our knowledge, almost nothing is inherently accessible to it.

The eviction order applies even to the occupancy of our own mental states. For Williamson, ‘there is no central core of mental states’, central in the special sense that ‘[i]f S belongs to that core, then whenever one attends to the question one is in a position to know whether one is in S.’ (p. 93) Such a state S (or the occupancy thereof) would be a case of a luminous condition holding. A condition C is luminous just in case (p. 95)

[i]or every case \( \alpha \), if in \( \alpha \) C obtains, then in \( \alpha \) one is in a position to know that C obtains

—where

[i]if one is in a position to know \( p \), and one has done what one is in a position to do to decide whether \( p \) is true, then one does know \( p \)... if one is in a position to know \( p \), then \( p \) is true.

Williamson maintains that (‘with rather trivial exceptions’—p. 13) no proposition can be the content of luminous knowledge. That is, even though an agent \( A \) might know that \( p \), it would not follow from that fact alone that \( A \) would be in a position to know that \( A \) knows that \( p \). This claim is of
course no threat to the knowability principle of the anti-realist, who would maintain only, concerning the assumed truth ‘A knows that p’, that it is knowable—that is, that it could be known at some time by some agent, even if not by the agent A himself.

This essay ventures to offer a modest dose house to Williamson’s epistemic hobo. It will be argued that thinkers have cognitive homes with the following minimal chattels: when they are in a state of understanding, with respect to any sentence, that it is meaningful for them (as a representation of how things are), then they indeed know (or at least are in a position to know) that they are in that state. That is to say, the condition of having a given declarative sentence be meaningful for one is luminous.\(^1\) The case to be made here will not be taking issue with Williamson on anti-realist grounds. Rather, it will take issue with him by adducing a kind of knowledge on the part of any agent which, simply by virtue of his having it, must be knowledge that he knows he has. The knowledge in question is therefore luminous, in Williamson’s sense. And it is not subject to any kind of Sorites-based scepticism at higher order. It is knowledge to the effect that one grasps a given sentence as meaningful—as making a declarative statement. In order to pick up the trail to such knowledge, we have to turn to ‘the other place’.

2 Moore’s cognitive phenomenology

G. E. Moore, in his essay ‘Propositions’ ([3], pp. 52–71; at p. 57–9), wrote of ‘some act of consciousness which may be called the understanding of [words’] meaning.’ For Moore, it was ‘plain that the apprehension of the meaning of one sentence with one meaning, differs in some respect from the apprehension of another sentence with a different meaning.’

This meaning-apprehension was a form of cognitive phenomenology. It was likened, by Moore, to the phenomenology of sensory qualities. He could not, however, say whether the two kinds of phenomenology were the same, or distinct (pp. 67–8). For Moore, the claim of cognitive phenomenology was not a conclusion based on philosophical argument. It played the role, rather,

\(^1\)Note that this is a different claim of luminosity than the one that Williamson finds implicit in Dummett’s claim in [1] at p. 131, that

[i]t is an undeniable feature of the notion of meaning—obscure as that notion is—that meaning is transparent in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same.
of an intuitive *given*, from which he would proceed, later in his essay, to make
certain other points about propositions. Apprehensions of meanings were
simply members of ‘a new class of facts which I want to call your attention
to’ (p. 56). They were ‘quite plain’ (p. 58).

Moorean propositions are in one regard, at least, different from Fregean
thoughts. For Frege, thoughts are senses of sentences. A thought ‘gets
clothed in the perceptible garb of of a sentence, *and thereby we are enabled
to grasp it*’ (Frege [2], at pp. 4–5; emphasis added) whereas for Moore (p. 61)

> We may ... apprehend a proposition, which we desire to express,
before we are able to think of any sentence which would express
it. ... we ... often apprehend [propositions], when we neither
see nor hear any words which express them, and probably often
without even having before our minds any *images* of words which
would express them.

Moorean propositions, however, like Fregean thoughts, are public and ob-
jective. On p. 63 Moore tells his reader

> What we ... mean ... when we say that [two] persons have the
same belief, is that *what* is believed in both of the two different
acts is the same: we mean by a belief, in fact, *not* the act of
belief, but *what* is believed; and what is believed is just nothing
else than what I mean by a proposition.

So: the propositions are public and objective, while our apprehendings of
them may be private. But, it may be emphasized, so it is too with col-
ors: colors may well be public and objective, while the *color-qualia* in any
individual mind are nevertheless private.

In course of naturalizing philosophy since Moore wrote on these mat-
ters, and under the influence of linguistic behaviorism in particular, many
philosophers of language were inclined to dispense altogether with proposi-
tions as theoretical entities. For some peculiar reason, hardly any attention
was paid to the plain facts of cognitive phenomenology to which Moore
drew our attention. Yet a mistaken impression of something is no less real,
as a mental phenomenon, for its turning out to be an impression of a non-
existent. To the extent that we are reminded, today, of the ‘hard’ problem
of consciousness, it is a problem generated almost exclusively by the recal-
citrant nature of sensory qualia, which, some believe, still elude any kind of
physical, behavioral-dispositional, or functional reduction.

There is an historical irony here. Philosophers working within a broadly
Quinean tradition seem to be in the grip of an unarticulated dogma. This
time the dogma does not involve the analytic/synthetic distinction, but is based on another important Kantian dichotomy—that between the operation of our sensibility and the operation of our understanding. The dogma in question holds that recalcitrant qualia—the ones that make the problem of consciousness hard—are confined to the side of sensibility. This is the ‘phenomenal’ or ‘subjective’ side of mind. For one who holds the dogma, it is as though the operation of a Kantian understanding could be ‘functionalized away’ to general satisfaction. (The operations of Kantian understanding are, today, subsumed under what is variously called ‘intentional’ or ‘psychological’; while the word ‘phenomenal’ covers what has to do, by and large, with Kantian sensibility.)

By reminding oneself of Moore’s simple insight, one hopes to bring post-Quinean philosophers of mind to the realization that they might have been, all along, in the grip of yet another unexamined dogma—the dogma to the effect that what it is like for a knowing subject is a matter only of sensibility, with the understanding contributing nothing at all to the phenomenology involved. (This dogma is to be distinguished from the so-called ‘third dogma of empiricism’, to the effect that there is a scheme/content distinction.)

There are a few contemporary philosophers who have revived the Moorean insight. Galen Strawson [6], [7] and Charles Siewert [5] are significant examples. With the orthodox opposition well on the way to Quining even sensory qualia, however, the burden of proof, even concerning what ought to be matters of simple intuition, seems to have shifted. (Here ‘intuition’ is used in the sense of ‘reflective datum’, a relative ‘given’ for the pursuit of a theoretical inquiry, rather than in the Kantian sense.) Today it is no longer enough to appeal to intuition the way Moore did; instead, a new kind of argument is needed.

An argument is offered below, whose purpose is to establish that there is such a thing as what it is like to apprehend a declarative proposition. It will establish this conclusion by means of a suitable thought-experiment. The thought-experiment is intended to reveal how apprehension of a proposition can be lamentably, but understandably, absent. It can be absent even when there are well-formed forms of words being exchanged by agents who happen to be unaware of their expressing propositions, given the very use that the agents are making of them.
3 Sensory qualia

The sensory qualia so beloved of phenomenologically minded philosophers are sometimes described as ‘raw feels’, but without any specifically haptic connotation. Examples are: a red patch in one’s visual field; a brief musical note heard; a bitter after-taste on the back of one’s tongue; an acrid whiff; a sharp stabbing pain on the tip of an index finger; a sudden wave of queasiness as one is lunched around. Across all sensory modalities, a Kantian would reckon these qualia to the manifold of sensory intuition. They are the primary mental stuff from which knowledge of one’s own body and of the external world are to be forged. It is only with the application of concepts that raw feels are synthesized into experience in Kant’s sense, i.e. cognition. When we descry objects, events and causal interactions, we are cuognizing. But this requires the use of various concepts of the understanding: sortal or classificatory concepts, and the relational concept of cause itself.

But there is a ‘pre-conceptual’ sense in which we still want to say, departing from Kant, that we experience qualia. Formless or uncategorized though this kind of experience might be, there is nevertheless something that it is like to undergo it. It is perhaps our most peripheral but also most acute form of awareness. It is the outermost surface of our sentence, where the impacts of the world are at their most vivid. Cognition, said Kant, needs this vivid raw material; for concepts without intuitions (here, qualia)—are empty (B75).

Even though, correlative and metaphorically, intuitions without concepts are blind, there is still something it is like for the subject to experience a motley sum of simultaneous raw feels, even short of appending the famous ‘I think’ of B137. And that something that it is like is pre-conceptual. The quality of one’s experience is grounded in qualia that are antecedent to experience in Kant’s sense. And those qualia are experienced in the ordinary sense.

Against this, some might say that even the rawest kinds of experience need a little conceptual massaging. We do not see little red patches; rather, we get glimpses of a tomato. We do not have any awareness of, say, the sound of a major fifth; it will be more definite than that, something more like major-fifth-on-a-piano. We do not have bitter after-tastes except with the awareness of their being after, and because of, our chewing those moldy nuts (say). We do not have, as it were, unsourced acrid whiffs; rather, such whiffs present themselves as of over-roasted coffee beans. One does not have just that sharp stabbing pain on the tip of one’s index finger; rather, one feels it as caused by some pointed object pressing against it. This is an
argument for at least some ‘conceptual infusion’ of even the most primitive phenomenological materials.

But all that is needed, as the departure point for our current reflections, is the something-that-it-is-like to be a subject enjoying qualia-informed-by-concepts. Consider just the sight of a tomato, short of any perceptual judgment concerning the tomato. There is something that it is like to see that red tomato, short of being drawn into any of a myriad possible states of judgment, such as that the tomato in question is round and ripe, or that it is within reach, or whatever. The awareness that is involved in cognition is still, basically, sensory awareness.

The familiar kinds of qualia just considered are the simplest phenomenal constituents in the something-that-it-is-like to have sensory awareness.

4 Semantic qualia

The goal now is this: to show how one can say, with Moore, that sensory qualia (with or without conceptual infusion) are not the only qualia. There are also semantic qualia. That is what the rest of this essay seeks to establish. Just as a visual percept might be broken down into its component color-qualia, so too the cognitive-phenomenological grasp (or conscious, Moorean understanding) of a proposition can be broken down into its component semantic qualia. The semantic correlate of one’s sensory awareness of a red patch in one’s visual field is one’s cognitive awareness that the word ‘and’ (say) means and. And the semantic correlate of one’s conscious visual perception of a scene is one’s conscious, Moorean understanding of the proposition expressed by a declarative sentence.

There is something-that-it-is-like to make logically structured judgments about the external world. There is something-that-it-is-like to represent the world (in private thought, or in communicative utterances) as being thus-and-so. If this was what Kant was onto in saying that it must be possible for an ‘I think’ to accompany any representation, then it cannot be claimed that the view was original with Moore; especially if Kant meant to allow also that there could be subjects whose representations happen not to be so accompanied. The sentential semantic quale \( \varphi \) is the content that one is aware of judging to hold—what one takes oneself, consciously, to be thinking—when offering a representation of the world as making \( \varphi \) true. Sentential semantic qualia are internally complex, on the one hand, but possessed also of an integrative unity, on the other. Their logical structure does not, however, detract from their experiential unity or immediacy.
The something-that-it-is-like in question here has nothing to do with one’s sensory manifold (except only indirectly, in certain cases). Instead, it is a kind of reflective awareness of the content and point of any given exercise of one’s conceptual and representational capacities, when one makes an assertion or assumes a proposition hypothetically, for the sake of argument. The propositional contents may be empirical; but they could also be mathematical. And the mere fact that sensory imagery is not necessary for abstract mathematical thought shows that conscious grasp of mathematical contents cannot presuppose the having of phenomenal qualia. One wants to say that there is something that it is like to judge that φ and to be aware of what, exactly, so doing amounts to; and that this something-that-it-is-like is in principle distinct from phenomenal experience.

It is very difficult to put this higher-order intuition into words. With qualia of any kind, there is the threat of ineffability. The color spectrum, for example, seems continuous, so in some sense has to be beyond the fumbling, discrete combinatorics of words. (Diana Raffman [4] has made this point convincingly.) Words fail qualia, even though (if Frege was right) words have to be adequate unto any judgment based on qualia. To have the sorts of raw feels in question is to know what it is that perhaps cannot be put into words. One has to let the subject find out for herself. One cannot say to her ‘Look, I’ll clue you in as to what sort of quale I’m talking about, and which you, lamentably, have never enjoyed. It’s like this, you see... ’, whereupon one tells her a long story that does not induce her to have any such qualia (such as the qualia of boredom at your holding forth), but instead lamely attempts to make her aware, by proxy as it were, of what she might have been missing. There can be no substitute for the real thing; she is going to have to find out, directly, for herself. ‘Cerise? It’s like this!’—whereupon one flashes a card in front of her, of the precise color with which she has never been acquainted. Then, and only then, does she know.

This is the method of the missed-out, but now made-up-for, kind of experience. Different subjects can get some idea of each other’s type-experiential lacunae; and they can visit upon one another the awareness of what, precisely, it is that they have been missing out on. It works, however, only for sensory qualia. It does not work for semantic qualia.

The method that works for semantic qualia is somewhat different. Recall the ancient Greeks’ story about the music of the spheres. We cannot hear those celestial strains, because they are (supposedly) always there. But if the music of the spheres were to stop, we would be acutely conscious of its cessation. A more mundane analogy would be the noise of cicadas in the tropics. One becomes so used to their near-incessant chirping that one
ceases to be aware of it. Awareness of the background din returns only when they all stop—which they will do, in concert, whenever they are disturbed.

Our foreshadowed thought-experiment is designed to make one appreciate the fact that there is something that it is habitually like to make judgments that represent the world as being thus-and-so. But, because it has always been like that for each of us for as long as we can remember, the reader has to be made aware of what this special kind of 'it' is like. The reader has to be enabled to imagine what it would be like—or how it would not be like anything at all—if it were not at all like that any more.

To exploit another analogy: consider the curious phenomenon of blindsight. Subjects with lesions in their brains can apparently process visual information, but without any visual awareness of the objects in their environment. Phenomenologically, as it were, they are blind; but, given how they move around and avoid obstacles in their way, they appear not to be. The thought is that, in the blindsighted subject, visual information is somehow being processed in the usual way—so that the brain mediates between visual perception and motor action—but that the subject, for some reason to do with the lesions, does not enjoy the usually concomitant visual awareness.

I shall seek to make the following analogy with blindsight. There could be subjects whose effectively communicative use of declarative language is, for all that natural selection might 'see', as good as our own, but for whom there is nothing that it is like to be making representational judgments about the world. (They are, if you like, 'judgmental zombies'.) The only difference is that with the victim of blindsight, the blindness is attributed on the basis of his own sincere claims to be visually unaware. In the case of semantic qualia, one cannot have a 'semantically snuff' subject making similar claims while in the grip of an attack of not-enjoying-semantic-qualia. Instead, the task here is to convince the reader, by means of a thought-experiment, that these imagined subjects are indeed (or at least: could very well be) semantically snuff, but that they could be made to acquire a nose for what they have been missing. They can become aware of what they have been missing, thereby making what has been missed a reality. But the reader, who has never missed it, needs this thought experiment in order to appreciate that one will be aware of having semantic qualia only by being made aware of what it would be like not to have them. This is like the ancient Greeks appreciating the very presence of the music of the spheres only when it ceases. Or, rather, it is like their appreciating the possibility of the presence of the music by being asked to imagine what it might be like if, should the music really be there, constantly, in the background, it were suddenly to cease.

Before the promised thought-experiment can be described, the reader
needs an explanation and description of a certain kind of game.

5 The game, described asemantically

Let there be two players, $\alpha$ and $\beta$. There are also two ‘role hats’, labelled, respectively, 1 and 2. At any stage of the game, each player will have exactly one of these hats on her head. Let the 1-1 assignment of hats to players be called $R$. Thus if $R(1) = \alpha$ (whence $R(2) = \beta$), player $\alpha$ has hat 1 on her head, while player $\beta$ has hat 2 on hers. If the players were to exchange their hats, the new hat-assignment would be $R$. Hence $\overline{R} = R$.

The players play the game with respect to ‘play-charts’ and a ‘playing-field’. A playing-field is a collection of individuals, each with various ‘properties’ and perhaps also standing in various ‘relations’. These properties and relations will be represented by predicate-letters (or simply, predicates). If one wanted to make a ‘playing-field’ for the purposes of having friends over to play the game, one might use a collection of solid shapes of different colors and sizes. One-place predicates could then represent shapes and colors, and two-place predicates such relations as ‘is bigger than’, ‘is darker than’, etc. If $P$ is an $n$-place predicate, and $x_1, \ldots, x_n$ are individuals in the playing-field $M$, then we shall write $P(x_1, \ldots, x_n)$ when those individuals stand in the relation represented by $P$ in the playing-field $M$. The statement $P(x_1, \ldots, x_n)$ belongs to the formalized language in which the game theory is given. Note that this formalized language has not yet been called a metalanguage!—for there is, as yet, no ‘object language’ under consideration.

The play-charts to be used alongside the playing-field are constructed by an inductive process. The basic materials from which one constructs play-charts are tags $x, y, z \ldots$; predicate-letters; and the three numerals 0, 1 and 2. All these primitive constituents (and those to be introduced below) are assumed to be distinct. That is, there is no overlap among any pairs of categories of primitive constituents. This is a standard condition to impose upon inductively defined entities, in order to secure the uniqueness of their method of composition. It is not peculiar to the study of semantically interpretable syntactic complexes.

Note that tags, predicate-letters and numerals are abstract types, and that one has to be able, in general, to distinguish different occurrences of some same tag, predicate-letter or numeral. Our rules of formation for play-charts are accordingly very abstract. But actual realizations of play-charts, for use by players, will be made out of concrete inscriptions, on pieces of paper, or board. Thus what is referred to abstractly as a play-chart of type
$n \psi \xi$ ($n = 1$ or 2), where $\psi$ and $\xi$ are themselves play-charts, might well be inscribed as nested rectangles, say:

\[
\begin{array}{|c|c|}
\hline
n & \psi & \xi \\
\hline
\end{array}
\]

Since there are only two values for $n$, one might adopt a convention whereby for $n = 1$ we have the diagram

\[
\begin{array}{|c|c|}
\hline
\psi & \xi \\
\hline
\end{array}
\]

and for $n = 2$ we have the diagram

\[
\begin{array}{|c|c|}
\hline
\psi & \xi \\
\hline
\end{array}
\]

The graphical possibilities are endless; assume only that some convention for play-chart inscription is settled upon, and adhered to. One can then proceed with the abstract inductive definition of play-charts, however they might be inscribed for use in actual play.

First there are the basic play-charts. These are formed by concatenating one predicate letter and an appropriate number of tags (among which repetitions may be permitted). Thus, with the two-place predicate $L$, one could form such basic play-charts as $Lxy$, $Lxx$, etc. The tags in a basic play-chart are said to be untied.

Secondly there are the ‘composite’ play-charts. If $\varphi$ and $\psi$ are any two play-charts, the following will also be play-charts:

- $0\varphi$ (whose untied occurrences of tags will be those of $\varphi$);
- $1\varphi\psi$ and $2\varphi\psi$ (whose untied occurrences of tags will be those that are untied in $\varphi$ and those that are untied in $\psi$).

Thirdly there are the ‘tying’ play-charts. Let $\varphi$ be a play-chart with an untied occurrence of $x$. Then $1x\varphi$ will be a play-chart, whose untied occurrences of tags will be those of $\varphi$, except for the untied occurrences of the tag $x$ in $\varphi$, all of which occurrences are now said to be tied by the prefix $1x$. Similarly for $2$ in place of $1$.

A play-chart with no tags untied is called a starting play-chart.

A play of the game involves a starting play-chart $\varphi$ and a playing-field $M$. The two players start by tossing a coin. The winner of the toss chooses which hat ($1$ or $2$) to wear at the outset. That choice of course determines
the initial hat-assignment $R$ for the play of the game. Play consists of finitely many successive stages. Each stage is characterized by

(i) the play-chart $\psi$ currently attended to;

(ii) the current hat-assignment; and

(iii) which individuals in $M$ have thus far been labelled by which tags untied in $\psi$ (see the rules below).

Thus a stage (or state of play) against the background of $M$ can be represented as

$$[\psi, R, f],$$

where $f$ is a finite mapping of the untied tags of $\psi$ to individuals in $M$. The notation $f(x/\mathbf{x})$ will represent the assignment that results from $f$ by extending it or modifying it so that the tag $x$ is assigned the individual $\mathbf{x}$.

With a starting play-chart $\varphi$ (which has no untied tags) the initial state of play will be of the form

$$[\varphi, R, \emptyset].$$

where $\emptyset$ is the empty assignment. The rules of the game are as follows:

1. If the current state of play is of the form $[0\psi, R, f]$ then the players must advance to the state of play $[\psi, \overline{R}, f]$. They have no choice in the matter. (The effect of the prefix 0 is thus to make the players swap hats as player 1 and player 2.)

2. If the current state of play is of the form $[n\psi\theta, R, f]$ ($n = 1$ or 2), then $R(n)$ chooses which of the following states of play will indeed be the next one: $[\psi, R, f]$; or $[\theta, R, f]$.

3. If the current state of play is of the form $[nx\psi, R, f]$ ($n = 1$ or 2), then $R(n)$ chooses an individual $\mathbf{x}$ from the playing-field $M$, and the next state of play will be $[\psi, R, f(x/\mathbf{x})]$.

4. If the current state of play is of the form $[P(x_1, \ldots, x_n), R, f]$, then play stops. If $P(f(x_1), \ldots, f(x_n))$, then $R(1)$ has won; otherwise, $R(2)$ has won.

Given these constitutive rules for this two-person, zero-sum game, it is easy to define the notion

$$\mathcal{P}_M[\psi, R, f]$$
(the player who can win, in state of play \([\psi, R, f]\), against the background of playing-field \(M\)). Consider first the two cases of states of play in which the players do not exercise any choices. The degenerate case where play ends (in a state of play involving a basic play-chart) is easy: the player who can win is simply the player who does win. Thus

\[
\mathcal{P}_M[P(x_1, \ldots, x_n), R, f] = R(1) \text{ if } \mathcal{P}(f(x_1), \ldots f(x_n));
\]

and

\[
\mathcal{P}_M[P(x_1, \ldots, x_n), R, f] = R(2) \text{ if } \neg \mathcal{P}(f(x_1), \ldots f(x_n)).
\]

Another easy case is where the state of play is of the form \([0\psi, R, f]\). Since the players advance to the next state of play without exercising any choices, the person who can win immediately before the hat-swap is the one who can win immediately after the hat-swap. Thus

\[
\mathcal{P}_M[0\psi, R, f] = \mathcal{P}_M[\psi, R, f].
\]

Or, put another way,

\[
\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff } \mathcal{P}(\psi, R, f) = R(2)
\]

—whence also

\[
\mathcal{P}_M[0\psi, R, f] = R(2) \text{ iff } \mathcal{P}(\psi, R, f) = \neg R(1).
\]

Next consider states of play in which the players do exercise choices. First look at such states of play that involve composite play-charts. What happens in a state of play of the form \([n\psi \theta, R, f]\) \((n = 1 \text{ or } 2)\)? The player \(R(n)\) gets to choose which of \(\psi\) or \(\theta\) to have in the next state of play. That is, \(R(n)\) gets to determine whether the next state of play is \([\psi, R, f]\) or \([\theta, R, f]\). Now here is a simple conceptual point about being in a winning position: *when it is your turn to make a choice, you are in a winning position before your choice if and only if there is a choice open to you that puts you in a winning position after it*. Moreover, if (and only if) at any stage all your choices would put your opponent in a winning position, then at that stage your opponent is in a winning position. Thus

\[
\mathcal{P}_M[1\psi \theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1);
\]

and

\[
\mathcal{P}_M[1\psi \theta, R, f] = R(2) \text{ iff } \mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2).
\]
Likewise
\[ P_M[2\psi \theta, R, f] = R(2) \text{ iff either } P_M[\psi, R, f] = R(2) \text{ or } P_M[\theta, R, f] = R(2); \]
and
\[ P_M[2\psi \theta, R, f] = R(1) \text{ iff } P_M[\psi, R, f] = R(1) \text{ and } P_M[\theta, R, f] = R(1). \]

Finally, look at states of play involving tying play-charts. What happens in a state of play of the form \([n x \psi, R, f]\) \((n = 1 \text{ or } 2)\)? The player \(R(n)\) gets to choose an individual \(x\) from the playing-field, and to tag it as \(x\). Thus \(R(n)\) gets to determine the next state of play, which has to be of the form \([\psi, R, f(x/x)]\). The simple conceptual point reasserts itself: when it is your turn to make a choice, you are in a winning position before your choice if and only if there is a choice open to you that puts you in a winning position after it. Moreover, if (and only if) at any stage all your choices would put your opponent in a winning position, then at that stage your opponent is in a winning position. Thus
\[ P_M[1 x \psi, R, f] = R(1) \text{ iff for some } x \text{ in } M \ P_M[\psi, R, f(x/x)] = R(1); \]
and
\[ P_M[1 x \psi, R, f] = R(2) \text{ iff for every } x \text{ in } M \ P_M[\psi, R, f(x/x)] = R(2). \]

Likewise,
\[ P_M[2 x \psi, R, f] = R(2) \text{ iff for some } x \text{ in } M \ P_M[\psi, R, f(x/x)] = R(2); \]
and
\[ P_M[2 x \psi, R, f] = R(1) \text{ iff for every } x \text{ in } M \ P_M[\psi, R, f(x/x)] = R(1). \]

6 At every state of play, exactly one player is in a winning position

Lemma. For all distinct \(\alpha, \beta\) and for all 1-1 functions \(R : \{1, 2\} \rightarrow \{\alpha, \beta\}\) the inductively defined function \(P\) from states of play to \(\{\alpha, \beta\} \ (= \{R(1), R(2)\})\) is total.


Basis. On states of play involving basic play-charts, it is determinate whether or not the state of affairs \(P(f(x_1), \ldots f(x_n))\) obtains. So \(P\) is total
on these states of play.

Inductive Hypothesis. Assume that \( \mathcal{P} \) is total on all states of play involving simpler play-charts than the play-chart \( \varphi \) under consideration. We consider \( \varphi \) by cases, according to whether it is of the form (i) \( 0\psi \), (ii) \( 1\psi \theta \), (iii) \( 2\psi \theta \), (iv) \( 1.x\psi \) or (v) \( 2.x\psi \) (where \( \psi \) and \( \theta \) are of course simpler than \( \varphi \)).

Case (i): \( \varphi = 0\psi \). By definition,

\[
\mathcal{P}_M[0\psi, R, f] = \mathcal{P}_M[\psi, \overline{R}, f].
\]

By Inductive Hypothesis, the right-hand side exists; whence the left-hand side \( \mathcal{P}_M[0\psi, R, f] \) is well-defined.

Case (ii): \( \varphi = 1\psi \theta \). The definition of \( \mathcal{P} \) specifies that

\[
\mathcal{P}_M[1\psi \theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1).
\]

By Inductive Hypothesis,

either \( \mathcal{P}_M[\psi, R, f] = R(1) \) or \( \mathcal{P}_M[\psi, R, f] = R(2) \);

and

either \( \mathcal{P}_M[\theta, R, f] = R(1) \) or \( \mathcal{P}_M[\theta, R, f] = R(2) \).

This yields the following four cases to consider:

\[
\begin{align*}
\mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1); \\
\mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2); \\
\mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1); \\
\mathcal{P}_M[\psi, R, f] = R(2) \text{ and } \mathcal{P}_M[\theta, R, f] = R(2).
\end{align*}
\]

In the first three of these cases, it follows by definition that

\[
\mathcal{P}_M[1\psi \theta, R, f] = R(1).
\]

In the fourth case, it follows by definition that

\[
\mathcal{P}_M[1\psi \theta, R, f] = R(2).
\]

Hence \( \mathcal{P}_M[1\psi \theta, R, f] \) is well-defined.

Case (iii) is similar to case (ii).

Case (iv): \( \varphi = 1.x\psi \). The definition of \( \mathcal{P} \) specifies that

\[
\mathcal{P}_M[1.x\psi, R, f] = R(1) \text{ iff for some } x \text{ in } M \mathcal{P}_M[\psi, R, f(x/\pi)] = R(1);
\]
and
\[ P_M[1x\psi, R, f] = R(2) \text{ iff for every } x \text{ in } M \quad P_M[\psi, R, f(x/x)] = R(2). \]

Unlike Case (ii), one cannot give a finite proof by cases here, for there might be infinitely many individuals \( x \). Note, however, that by Inductive Hypothesis we have

for every \( x \) in \( M \) either \( P[\psi, R, f(x/x)] = R(1) \) or \( P[\psi, R, f(x/x)] = R(2) \).

This is of the form \( \forall x(Ax \lor Bx) \). From this it follows by classical logic that \( \exists x Ax \lor \forall x Bx \). But the first disjunct is the condition for

\[ P_M[1x\psi, R, f] = R(1) \]

to hold; while the second disjunct is the condition for

\[ P_M[1x\psi, R, f] = R(2) \]

to hold. It follows that \( P_M[1x\psi, R, f] \) is well-defined.

Case (v) is similar to case (iv).

QED

7 The thought-experiment

Imagine a society in which there evolves a battle of wits among people of leisure. They came to play the foregoing game, and with great ceremony. That their elaborate rituals count as playing this game, under the abstract description given above, is the product of both anthropological and mathematical insight; for the matter is not at all that clear to the contestants involved.

Their playing-field is, literally, a playing-field. Figurines and other familiar objects can be displayed and re-located on the arena. There is a stock of monadic and dyadic predicates that determinately apply, or fail to apply, by universal agreement, to any object, or pair of objects, within the arena.

A coin-toss determines which player begins as Player 1, and which as Player 2. (The winner of the toss chooses.)

The player whose move it is makes a choice: either of a sub-chart, \( \psi \) or \( \theta \); or of an individual on the arena to be labelled with the untied tag \( x \). In the former case, the rest of the chart is covered, and only the chosen sub-chart remains exposed to view. In the latter case, where an individual
has to be chosen, the chooser places against it a placard inscribed with the appropriate tag. Recall that on a play-chart of the form $nx\psi$, (he who at that stage is) Player $n$ will choose some individual and tag it with $x$. The most able players will also remove tags from the objects in the tableau, as and when their presence would no longer be required, given the course of play that has transpired.

The professional circuit is dominated by the most agile minds. These are the geniuses who, after brief inspection of a fresh play-chart and a newly arranged playing field, can elect to be Player 1 or Player 2 at the outset, and display amazing resourcefulness in subsequently winning play after play of the game.

Small children are reared on the game in parlour form. Famous play-charts of yore can be purchased in miniature form at the local shops. For those who cannot afford the ersatz playing-fields with plastic figurines, there are booklets containing diagrammatic representations of famous playing-fields, indicating the layouts of the actual figurines and other objects to which they bore observable and determinate relationships.

In due course, in response to the needs of post-game analysis, someone develops a linear notation for describing play-charts. It is some peripatetic Pole who first thinks up the new notation. He writes $1\psi\xi$ for the two-dimensional play-chart

\[
\begin{array}{|c|c|}
\hline
\psi & \xi \\
\hline
\end{array}
\]

and similarly writes $2\psi\xi$ for the play-chart

\[
\begin{array}{|c|c|}
\hline
\psi & \xi \\
\hline
\end{array}
\]

Indeed, the Pole hits on exactly the notation that was used above for the inductive definition of play-charts! With a little practice, his interlocutors catch on to what these linear notations denote. They stand for the two-dimensional play-charts with which players would actually play at the big meets.

The Pole also realizes that the playing-fields need shorthand designations, but cannot devise a succinct naming system in general. He adopts, instead, the local nomenclature, involving short descriptive terms, such as ‘The Sicilian’, for playing fields that are really quite complex. Fortunately, his interlocutors are such avid followers of the game that they can recall
every structural detail about playing-fields thus described. In due course they came to say things like ‘So-and-so would have loved to be Player 1 on The Sicilian with 1\textit{x}2\textit{y}010\textit{LxyLyz’}.

Travelling home from work, colleagues might look at the local playing-field (on which the figurines were always left in their final arrangement at the end of any big meet), and venture tense comments like ‘2\textit{x}1\textit{y}2\textit{z}02\textit{Lxy}0\textit{Lzy’}. These were understood as meaning that the champion they had just been discussing would have been unbeatable as Player 1 on that playing-field, with the play-chart just designated. Such ‘comments’ would never be understood as assertions. If any ‘speech-act’ had to be read into such an utterance, it would be something like a performative: an expression of preparedness to bet that (whoever started as) Player 1 would win on the play-chart in question, on the playing-field in question.

But such opinions about how the great might wish to play are understood only as that: opinions about how the great might wish to play. The play-charts are \textit{mentioned}, not used, by devotees of the game. And even the great players who really do use them (by playing on them) do not use them in the sense relevant to the use-mention distinction in semantics. They use them, rather, as a chess-player uses the chess-board and the chess-pieces on it: that is, as non-semantic items by means of which one plays a game. When one looks at a board-position near the end-game, and says ‘White should win’, one does not take the board-position to be saying anything. For no \textit{p} does it bear the content that \textit{p}. In the same way, it would never occur to the devotees of our game that a play-chart said anything about the playing-field—that it bore a semantic content, determined by its structure, and was accordingly \textit{true} or \textit{false} as a representation of the playing-field. In relation to the play-charts, the gamers are without what Moore called ‘some act of consciousness which may be called the understanding of their meaning’.

8 The most important biconditionals; and a homomorphism

We leave our thought-experiment for the time being, and return to some game theory. It has a lesson for us about semantic qualia, and about how light can dawn suddenly over the whole.

Equipped with the Lemma proved in \S6, one can now deduce certain consequences of the definitional clauses specifying the notion $\mathcal{P}_M[\psi; R, f]$ above.
Making the classical assumption that it is determinate whether or not any basic fact obtains, one can infer:

$$\mathcal{P}_M[P(x_1, \ldots, x_n), R, f] = R(1) \text{ if and only if } P(f(x_1), \ldots f(x_n))$$

Because $\mathcal{P}$ is total, the clause for hat-swaps implies:

$$\mathcal{P}_M[\psi, R, f] = R(1) \text{ iff } \neg[\mathcal{P}_M[\psi, R, f] = \neg R(1)]$$

Recall also that we already have the following:

$$\mathcal{P}_M[\psi \theta, R, f] = R(1) \text{ iff either } \mathcal{P}_M[\psi, R, f] = R(1) \text{ or } \mathcal{P}_M[\theta, R, f] = R(1)$$

$$\mathcal{P}_M[2\psi \theta, R, f] = R(1) \text{ iff } \mathcal{P}_M[\psi, R, f] = R(1) \text{ and } \mathcal{P}_M[\theta, R, f] = R(1)$$

$$\mathcal{P}_M[1x \psi, R, f] = R(1) \text{ iff for some } x \text{ in } M \mathcal{P}_M[\psi, R, f(x/x)] = R(1)$$

$$\mathcal{P}_M[2x \psi, R, f] = R(1) \text{ iff for every } x \text{ in } M \mathcal{P}_M[\psi, R, f(x/x)] = R(1)$$

Thus far no attempt has been made to interpret play-charts as making statements about the playing-field. Let $\varphi$ be a play-chart and let $f$ assign to each tag untied in $\varphi$ either an individual (from $M$) or a variable in our formalized language for game theory. No two distinct tags may be assigned the same such variable by $f$.\(^2\) Suppose $f$ is an assignment dealing with exactly the untied tags in $\varphi$, and suppose further that $\psi$ is a subformula of $\varphi$. Then $f^\psi$ will be the restriction of $f$ to tags that are free in $\psi$. Now define a homomorphic mapping $\tau_M^f$ from play-charts $\varphi$, and relative to such

\(^2\)Strictly speaking, since this discussion of $f$ and $\tau_M^f$ is taking place in a language one level up from the language of our game theory, one should use, as the mapping referred to by this superscript, not $f$ itself, but a mapping $f^*$ closely related to $f$. Where $f$ assigned to any tag an individual from $M$, $f^*$ would assign to that tag some name (in the language of our game theory) for the individual. This would ensure that $\tau_M^f(\varphi)$ would always be a sentence of the language of our game theory. Here, however, one simplifies slightly by avoiding such niceties. Alternatively, one could just stipulate that the language of our game theory simply contained all the relevant expressions of its own metalanguage.
assignments \( f \), to sentences of our formalized language of game theory.

\[
\tau^f_M(P(x_1, \ldots, x_n)) = df \ P(\tau^f_M(x_1), \ldots, \tau^f_M(x_n));
\]

\[
\tau^f_M(0\psi) = df \ -\tau^f_M(\psi);
\]

\[
\tau^f_M(1\psi\theta) = df \ (\tau^f_M(\psi) \lor \tau^f_M(\theta));
\]

\[
\tau^f_M(2\psi\theta) = df \ (\tau^f_M(\psi) \land \tau^f_M(\theta));
\]

\[
\tau^f_M(1x\psi) = df \ \exists x\tau^f_M(x/\psi);
\]

\[
\tau^f_M(2x\psi) = df \ \forall x\tau^f_M(x/\psi).
\]

Note that nothing more than a syntactic mapping has been defined here, from play-charts to sentences of the formalized language in which the game theory is being presented.

Here is an example to illustrate the action of \( \tau^f_M \) as just defined. Consider the play-chart

\[ 2y10Ayx1xBx. \]

Note that the tag \( x \) has both tied and untied occurrences therein. The leftmost occurrence of \( x \) is untied. No other tag has an untied occurrence. So suppose \( f \) maps \( x \) to the individual \( \gamma \) in \( M \). Then \( f \) is simply \( \{<x, \gamma>\} \).

We can now calculate \( \tau^f_M\{<x, \gamma>\}(2y10Ayx1xBx) \) as follows:

\[
\tau^f_M\{<x, \gamma>\}(2y10Ayx1xBx)
\]

\[
= \forall y \tau^f_M\{<x, \gamma>, y, y'>\}(10Ayx1xBx)
\]

\[
= \forall y (\tau^f_M\{<x, \gamma>, y, y'>\}(0Ayx) \lor \tau^f_M(1xBx))
\]

\[
= \forall y (\neg \tau^f_M\{<x, \gamma>, y, y'>\}(Ay) \lor \exists x \tau^f_M\{<x, \gamma>, y, y'>\}(Bx))
\]

\[
= \forall y (\neg \exists x (\tau^f_M\{<x, \gamma>, y, y'>\}(x), \tau^f_M\{<x, \gamma>, y, y'>\}(y), \exists x B(\tau^f_M\{<x, \gamma>, y, y'>\}(x)))
\]

\[
= \forall y (\neg A(y, \gamma) \lor \exists x B(x)).
\]
9 Towards Convention T

Theorem. Suppose \( f \) deals with all the untied tags of \( \varphi \). Then for all \( R \),
\( \mathcal{P}_M[\varphi, R, f] = R(1) \) is interdeducible with \( \tau_M^f(\varphi) \).

Proof. By induction on the complexity of the play-chart \( \varphi \).

Basis. Immediate from the basis clause in the definition of \( \tau_M^f \) and the first boxed biconditional above.

Inductive Hypothesis. Suppose the result holds for all play-charts simpler than \( \varphi \).

Inductive Step. Consider \( \varphi \) by cases. We shall deal here only with the case
where \( \varphi \) is of the form \( 0\psi \). We have

\[
\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff not-} [\mathcal{P}_M[\psi, R, f] = R(1)].
\]

By Inductive Hypothesis, we obtain

\[
\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff not-} [\tau_M^f(\psi)].
\]

By definition of \( \tau_M^f \), it follows that

\[
\mathcal{P}_M[0\psi, R, f] = R(1) \text{ iff } \tau_M^f(0\psi).
\]

The remaining cases are similar.

Corollary. If \( \varphi \) is a starting play-chart, then for all \( R \), \( \mathcal{P}_M[\varphi, R, \emptyset] = R(1) \)
is interdeducible with \( \tau_M^f(\varphi, \emptyset) \).

Note that \( \tau_M^f(\varphi) \) does not involve \( R \). It follows that if any

\[
R : \{1, 2\} \xrightarrow{1-1} \{\alpha, \beta\}
\]

makes it the case that \( \mathcal{P}[\varphi, R, f] = R(1) \), then every such \( R \) makes it so.

Our theorem above therefore essentially correlates a relational property—
call it \( S \)—of \( \varphi \), \( M \) and \( f \) with the condition \( \tau_M^f(\varphi) \). And in the special case
where \( \varphi \) is a starting play-chart (with all tags tied) this relational property
\( S \) of \( \varphi \), \( M \) and the null assignment is being correlated with the condition
\( \tau_M^f(\varphi, \emptyset) \). The latter is simply a statement about the model \( M \). One can
further ignore the null assignment, and write \( T(\varphi, M) \) instead of \( S(\varphi, M, \emptyset) \);
and write \( \tau_M(\varphi) \) for \( \tau_M^0(\varphi) \).

One has therefore fulfilled a model-relative form of Tarski’s material
adequacy condition on definitions of would-be truth-predicates:
for every play-chart \( \varphi \) with no tags tied, and for every playing-field \( M \), \( T(\varphi, M) \) is interdeducible with \( \tau_M(\varphi) \).

Now Tarski took the view that if \( \tau_M \) was a translation mapping relative to \( M \) from what were regarded as object-language sentences into (what would accordingly be regarded as) the metalanguage, then fulfillment of his material adequacy condition ensured that \( T(\varphi, M) \) would capture the truth of \( \varphi \) in \( M \).

No claim has yet been entered to the effect that these ‘play-charts’ are sentences—that is, interpretable as statements about the ‘playing-field’ \( M \). The charts have been treated as non-semantic, albeit quasi-syntactic structures.

It is time now to take a leap. Let the play-charts be interpreted in the obvious way given by \( \tau \) above: basic play-charts are atomic predications; 0 is negation; 1 is disjunction; 2 is conjunction; 1\( x \) is the existential quantifier; and 2\( x \) is the universal quantifier. The reader will be quick to realize that play-charts, as they were presented above, are in Polish notation. Tags are variables; the tied and untied ones are bound and free, respectively. Everything now falls into place. Courtesy of Tarski’s adequacy condition on theories of truth, one can now regard the game-theoretical property

\[
P_M[\varphi, R, 0] = R(1)
\]

as that of the truth of \( \varphi \) in \( M \).

A mathematically sophisticated member of the culture in our thought-experiment could be brought, by the foregoing considerations, to realize that play-charts are propositional representations, and that they represent the ways things are on the playing-field in question. They will be able to ‘flash-grasp’ that

1. Player 1 is asserting the proposition expressed by the playing-chart;
2. Player 2 is denying it;
3. Player 1 can win, regardless of what Player 2 might do, provided that the playing-chart is true of the playing-field;
4. Player 2 can win, regardless of what Player 1 might do, provided that the playing-chart is false of the playing-field;
5. 0 means negation;
6. 1 means inclusive disjunction;
7. 2 means conjunction;

8. 1x means ‘for some x’;

9. 2x means ‘for every x’;

10. play-chart construction is in accordance with grammatical rules;

11. truth and falsity are objective;

12. a winning strategy for Player 1 on a (finite) playing-field takes the form of an ‘evaluation proof’ of the play-chart with respect to the field;

13. a winning strategy for Player 2 on a (finite) playing-field takes the form of an ‘evaluation disproof’ of the play-chart with respect to the field;

14. the play-chart’s truth-conditions can be expressed by a sentence (in a logician’s version)\(^3\) of one’s ordinary language (assuming that it contains equivalents of the aforementioned logical operators, along with the necessary extralogical predicates)

The point is, all these insights can dawn in one sudden revelation, enlightening the gamer as to the true nature of what he is up to. Coming to see any given play-chart as truth-evaluable (on any given playing-field) is a matter of seeing that it is true if and only if … (where the dots are filled in by some translation of the play-chart into the expressively adequate language, or logician’s version thereof, that the formerly naïve gamer has already mastered).

Before such enlightenment, the gamer has no semantic qualia in connection with the play-charts, against the background of a playing-field. After his enlightenment, the gamer cannot help but experience the play-chart as semantically contentful, as making a truth-evaluable claim about the playing-field, a claim that is translatable into (a slightly formalized extension of) one’s ordinary language. The gamer formerly used descriptions of play-charts (descriptions constructed in what he took to be an extension of his ordinary language), and he uttered them in contexts in which his speech-act was at best intended as a performative. He now realizes that these erstwhile descriptions can be re-interpreted as declarative sentences in their own right.

\(^3\)What is meant here is a language which, like so-called ‘logician’s English’, allows one to render bound variables explicitly in surface form, rather than attempt to deal with them only by means of anaphoric pronouns.
They form a new language. He realizes that he is, in effect, using these sentences to make assertions about the playing-field. He is now capable of that Moorean ‘act of consciousness which may be called the understanding of their meaning.’ A certain intellectual light has dawned on the whole. The sentences in question now afford the underlander semantic qualia. He knows that the sentences are meaningful representations; and, by virtue of that fact alone, he knows that he knows it. Semantic knowledge is luminous.

References


