Signaling systems and the Transcendental Deduction

(Forthcoming in a collection on Idealism edited by Tyron Goldschmit and Kenny Pearce)

The central claim of Kant’s Transcendental Deduction is that the unity of consciousness entails the objectivity of experience. I don’t know exactly what he meant by that, why he thought it was true or what that has to do with the sort of idealism he opposed. But the claim obviously has great interest; and I do think it is both true and interesting that feeding it through a positivist meat-grinder, as is done here, yields something that is recognizably a version of *it*, but has nothing especially to do with imagination, or thought, or language, or the categories. It is a general truth about arbitrary signaling systems.

That truth may nonetheless cast some light on (a) Kant and (b) people. Sections 3 and 4 at least gesture at that; whilst section 2 covers the model that my main argument concerns. The illustrative story in section 1 is meant to give an intuitive feel for the key premise (18).

1. The story concerns a little-known episode from the last days of the Trojan War.[[1]](#footnote--1) Everyone knows that the Greeks left outside the gates a wooden horse stuffed with soldiers. Few know that some Trojans set up communication with the horse. The language *L* in which they did this, the contemporary lingua franca amongst wooden horses, is coincidentally very like a primitive fragment of English. Its spoken *vocabulary* has four categories:

1. *Colour* words (‘red’, ‘green’, ‘blue’ etc.)
2. *Shape* words (‘circle’, ‘square’, ‘triangle’ etc.)
3. Signs for *affirmation and negation* (‘yes’, ‘no’)
4. Inflective markers of *assertion and interrogation* (transcribed ‘!’, ‘?’)

The *sentential utterances* of L come in three categories, of which these two are initially relevant:

1. *Simple observational utterances*: these may be (i) assertions (‘Red!’, ‘Green!’, ‘Square!’ etc.) or (ii) questions (‘Red?’, ‘Green?’, ‘Square?’ etc.)
2. *Sentences of assent and dissent*: ‘Yes!’, ‘No!’

Although the horse could only *utter* sentences in class (6) (‘Yes!’ and ‘No!’), it could *understand* simple observational questions i.e. from class (5ii). If a Trojan asked ‘Red?’ the horse would say ‘Yes!’ when, and only when, something *red* lay in its line of sight, otherwise ‘No!’; similarly for all questions in class (5ii).

Everyone wondered what was going on *inside* the horse. Most agreed that there were *people* in its head looking through its eyeholes. There were two hypotheses:

1. The **fragmentary hypothesis**: there are *two* people in the head: someone with blurred vision who sees colours but not shapes, and someone with black-and-white vision who sees shapes but not colours. When asked a question they consult before one of them answers through the loudspeaker in the horse’s mouth.
2. The **unitary hypothesis**: *one* (visually unimpaired) person is looking through an eyehole. When asked, he answers.

No amount of questioning could settle the issue between (7) and (8).

At least not until someone remembered a third category of utterance in *L*:

1. *Compound observational utterances*: these concatenate a word from class (1) and a word from class (2); they may be (i) assertions (‘Blue square!’, ‘Red circle!’ etc.) or (ii) questions (‘Blue square?’, ‘Red circle?’ etc.)

Now it was easy. One simply asked a series of *compound* questions: if the horse can answer them correctly then that rules out (7); if not, it rules out (8). Here, a correct answer means e.g. that the horse says ‘Yes!’ to ‘Red triangle?’ if and only if something that is visibly *both* red and triangular lies in its line of sight.

This works because (8) implies that the horse can distinguish (i) a red triangle and a blue square from (ii) a blue triangle and a red square. (7) implies that it cannot do this, even if the people in its head can consult. The Trojans discovered that *two* people were in the horse’s head. They should then have started to wonder what they were doing there—but they didn’t, and we needn’t.

*2. Unity and objectivity*

The formal exposition begins with the very simple model to which I am going to apply it. Complicating things to make it more realistic is obviously desirable, but doing it here would (i) take more space than I have and (ii) obscure the main lines of argument. Subsection 2.1 sets out and briefly motivates the elements of the model by means of twelve definitions and stipulations. 2.2 establishes the main result concerning the model.

*2.1 Formal framework*.

(a) We assume a set of ***objects*** *o*1, *o*2… *om* and a set of monadic ***properties*** *f1*… *fn* that might be distributed over those objects in any of various ways.[[2]](#footnote-0)

(b) The variety of these ways gives rise to a set S of possible ***states of the world***, each one being a specification, for each property, of its extension i.e. of what has it. So given *n* objects and *m* properties there are at most 2*mn* states. ‘At most’ because there may be combinatorial restrictions: for instance, it may be impossible for *f*1 and *f*2 to apply to one object. But at this point I’ll ignore that kind of restriction: allowing for it complicates matters but seems to make no other difference.

An example with *m* = *n* = 2: a screen is divided into left and right and programmed to display at regular intervals a shape (triangle or other) in a colour (red or other) on one side, both sides or neither side e.g. a red square on the left and a blue triangle on the right; or a blue square on the right and nothing on the left. *o*1 and *o*2 are the left and right sides, and *f*1 and *f*2 might be the properties of displaying something red and something triangular respectively. Sixteen possible states of the world correspond to the possible assignments of subsets of {*o*1, *o*2} to *f*1 and *f*2:

|  |  |  |
| --- | --- | --- |
| **State** | **Assignment to *f1*** | **Assignment to *f2*** |
| *s*1 | ∅ | ∅ |
| *s*2 | ∅ | {*o*1} |
| *s*3 | ∅ | {*o*2} |
| *s*4 | ∅ | {*o*1, *o*2} |
| *s*5 | {*o*1} | ∅ |
| *s*6 | {*o*1} | {*o*1} |
| *s*7 | {*o*1} | {*o*2} |
| *s*8 | {*o*1} | {*o*1, *o*2} |
| *s*9 | {*o*2} | ∅ |
| *s*10 | {*o*2} | {*o*1} |
| *s*11 | {*o*2} | {*o*2} |
| *s*12 | {*o*2} | {*o*1, *o*2} |
| *s*13 | {*o*1, *o*2} | ∅ |
| *s*14 | {*o*1, *o*2} | {*o*1} |
| *s*15 | {*o*1, *o*2} | {*o*2} |
| *s*16 | {*o*1, *o*2} | {*o*1, *o*2} |

*Table 1: States of the world (∅ the empty set)*

E.g. *s*1 corresponds to cases where nothing is red or triangular; *s*8 to those with a red triangle on the left and a non-red triangle on the right; *s*14 to those with a red triangle is on the left and a red non-triangle on the right; and *s*16 corresponds to the case in which both shapes are red triangles.

We associate with each property a function from S to the power set of {*o*1, … *om*} that takes each *s* to *f*’s extension in *s*. I will use *f* itself to denote this function as well as the property; thus concerning Table 1 we have e.g. *f*1 (s1) = Ø, *f*2 (*s*15) = {*o*2}.

(c) A ***signal*** is a function *p* from S to {0, 1}. Informally, think of each signal as a question to which the answer in a state is ‘yes’ if the output at that state is 1 and ‘no’ if it is 0. In this example, corresponding to the question ‘Is something red?’ is the signal *p*1 that takes 1 at exactly *s*5 through *s*16 i.e. states at which *f*1 is not empty. I’ll write ***[p]*** for the set of states *s* such that *p* (*s*) = 1.

(d) If *f* is a property and *p* a signal then *p* is an *{f}-signal* if its value supervenes on the extension of *f*. If any two states agree over what is *f* (whether or not over anything else), they agree over *p*. Formally and with an obvious extension:

1. *p* is an ***{f}-signal*** if ∀*s*∀*s*\* (*f* (*s*) = *f* (*s*\*) → *p* (*s*) = *p* (*s*\*)). It is a ***non-trivial {f}-signal*** if in addition ∃*s*∃*s*\* (*p* (*s*) ≠ *p* (*s*\*)).

*p*1 is thus an {*f*1}-signal, as is that corresponding to ‘Is *o*1 red?’ But *p*2, corresponding to the question ‘Is something *triangular*?’, is *not* an {*f1*}-signal: *s*5 and *s*6, for instance, agree over which objects are *f*1 (i.e. exactly *o*1) but *p*2 (*s*5) = 0 and *p*2 (*s*6) = 1. On the other hand *p*2 *is* an {*f*2}-signal.

(e) Next the obvious generalization: for an arbitrary set *F*of properties, *p* is an *F-signal* if its value supervenes on the extensions of the *F*s. Formally:

1. For any set *F* of properties, *p* is an ***F-signal*** if ∀s∀s\*(∀*f* ∈ **F** (*f* (*s*) = *f* (*s*\*)) → *p* (*s*) = *p* (*s*\*)).

Table 1 doesn’t readily illustrate this: *all* signals are {*f1*, *f2*}-signals. But if we have a set *F* of chromatic properties (*blue*, *red*, *green*…) and a set *G* of shape properties (*triangular*, *square*, *circular*…) then the signal that takes value 1 at worlds where *something is green and something is red* is an *F*-signal but not a *G*-signal; that which takes value 1 at states where *something is green and something is triangular* is *neither* an *F*-signal nor a *G*-signal, but it *is* an F ∪ G-signal.

Note that if *F* ⊆ *G* then all *F*-signals are *G*-signals but some *G*-signals may not be *F*-signals; in fact a *G*-signal may not be an *F*-signal for *any* *F* ⊂ *G*. Table 1 does illustrate this. Let *p*3 (*s*) = 1 just in case something is *both* *f*1 and *f*2 at *s*: so [*p*3] = {*s*6, *s*8, *s*11, *s*12, *s*14, *s*15, *s*16}. *p*3 is an {*f*1, *f*2}-signal but not an {*f*1}-signal or an {*f*2}-signal.

(f) A ***signaling system P***is a set of signals. E.g. in connection with Table 1, the signaling system {*p1*, *p*2} corresponds to a person seated before the computer screen who is trained only to answer ‘Is something red?’ and ‘Is something triangular?’ The values of *p1* and *p*2 correspond to his dispositions to answer each question ‘Yes!’ or ‘No!’ But language and consciousness are both inessential: what matters is only the correlation between the outputs of the signaling system and the state of the world. {*p1*, *p*2} equally corresponds to a machine that flashes a red light when something on the screen is red and a green light when something is triangular.

(g) Consider that probably fictitious stage of linguistic development at which the learner only registers the *presence* *or* *absence* of something in his environment having a property to which he is sensitive. ‘Red!’ or the disposition to assent to it occurs when some local object or region is red, dissent when none is; similarly ‘Rabbit!’ when there is a rabbit around, and so on. Fictitious or not we can certainly *imagine* children, animals or artificial devices that are in it: and they are *object*-*insensitive*.

They can be modeled as signaling systems of the sort described in our definitions, given a suitable ontology. Let *f* apply to something if it is a visible region of space and white; let *f*\* apply to something if it is a local rabbit; and so on. And let the ontology include eligible candidates for these properties, for instance rabbits and other physical objects as well as appropriately sized regions of space. Then at any particular time the linguistic dispositions of a signaling system of this sort correspond to its signals. And across those possible worlds (within a suitably realistic range) at which the properties are distributed over the objects, these signals supervene on that distribution.

Formally, object-insensitivity is defined as follows:

1. A signaling system *P* is ***object***-***insensitive*** if for any set *F* of properties, for any *F*-signal *p* ∈ *P*, ∀*s*∀*s*\*((∀*f* ∈ *F* (*f* (*s*) = Ø ↔ *f* (*s*\*) = Ø) → *p* (*s*) = *p* (s\*)). Otherwise *P* is ***object-sensitive***.

What (12) means is that if two states of the world agree on whether (e.g.) *something* is white and on whether *something* is a rabbit – never mind what – then an object-insensitive system has the same output at both states, at least with regard to those signals that are sensitive only to the distribution of whiteness and/or rabbits. It could not distinguish states that agree that *something* is white and *something* is a rabbit but differ over which.

It is clear why I call this object-insensitivity. An object-insensitive system may register the instantiation of *f* but not *which* object it is, how many things have it or where one *f* ends and another begins. Its {*f*}-signals register only whether there is *f*-ing going on. In this respect they behave like what Evans calls feature-placing expressions – like ‘Lo, rabbithood!’ in the tongue of one who can neither distinguish nor identify rabbits.[[3]](#footnote-1) Similarly, if an object-insensitive system is sensitive to *f* *and* *f*\*, then it doesn’t care whether any *f* is the same as or different from any *f*\* but only whether any *f*-ing and/or *f*\*-ing is going on.[[4]](#footnote-2)

(h) If *f* and *f*\* are properties and *P* is a signaling system then I’ll say *P separates* *f* and *f* if all of P’s {*f*, *f*\*}-signals supervene on its {*f*}-signals and its {*f*\*}-signals. That is: if any signal in *P* gives the same value at states that agree on the extensions of *f* and *f*\* then it gives the same value at states that agree on the values of its {*f*}-signals and also on the values of its {*f*\*}-signals

1. ***P separates f and f\**** if ∀*p* ∈ *P* (*p* is an {*f*, *f*\*}-signal → ∃*Q* ⊆ *P* (∀*q* ∈ *Q* (*q* is an {*f*}-signal ∨ *q* is an {*f*\*}-signal) ∧ ∀*s*∀*s*\* (∀*q* ∈ *Q* (*q* (*s*) = *q* (*s*\*)) → *p* (*s*) = *p* (*s*\*))).

Thus e.g. in Table 1, let *p*4 (*s*) = 1 iff something is *f*1 at *s* *and* something (possibly something else) is *f*2 i.e. [*p*4] = [*p*1] ∩ [*p*2]. Then the signaling system {*p*1, *p*2, *p*4} separates *f*1 and *f*2, since *all* its signals take the same value at states where the {*f1*}-signal *p*1 and the {*f*2}-signal *p*2 take the same value. On the other hand {*p*1, *p*2, *p*3} does *not* separate *f*1 and *f*2 since *p*3 – defined as above as indicating that some *one* thing is *f*1 and *f*2 – takes value 1 in some states where *p*1 = *p*2 = 1 (e.g. *s*6) but 0 at other such states (e.g. *s*12).

(j) If *P* is a signaling system then two states are *P-indiscernible* if all signals in *P* take the same values across them:

1. *s* and *s*\* are ***P-indiscernible***, written ***s ~P s\****, if ∀*p* ∈ *P* (*p* (*s*) = *p* (*s*\*))

For instance, the signaling system {*p1*, *p*2} cannot discern by its output two states in which something is red and something is triangular, since in all such cases it answers ‘Yes’ to both questions: any two such states are therefore {*p*1, *p*2}-indiscernible. ~{*p1*, *p*2} thus generates four equivalence classes: {*s*1} (nothing red or triangular), {*s*2, *s*3, *s*4} (nothing red but something triangular), {*s*5, *s*9, *s*13} (something red but nothing triangular) and {*s*6, *s*7, *s*8, *s*10, *s*11, *s*12, *s*14, *s*15, *s*16} (something red and something triangular).

(k) Two signaling systems *P* and *P\** are ***operationally equivalent*** if ~*P* = ~*P*\* i.e. states are *P*-indiscernible if and only if they are *P\**-indiscernible. This equivalence relation between signaling systems is not identity: a signaling system with three signals might be operationally equivalent to one with two. For instance, if we define *p*4 as in 2.1(h) then {*p1*, *p*2} and {*p1*, *p*2, *p*4} are operationally equivalent. In fact if [*p*] results from Boolean operations on [*p*1] and [*p*2] then {*p*, *p1*, *p*2} and {*p*1, *p*2} are operationally equivalent.

Informally, the idea behind operational equivalence is that operationally equivalent signaling systems make the same discriminations. The totality of answers from someone that we could only ask whether something is red and (separately) whether something is triangular, would tell us as much about his environment as the answers from somebody who could also answer the *conjunctive* (not compound) question ‘Is there something red *and* something triangular?’

(l) Again an obvious generalization: a signaling system *P* and a *set* Π of signaling systems are operationally equivalent if the *totality* of signals from signaling systems in Π carries just as much information as the totality of signals from *P*.

1. If *P* is a signaling system and Π is a *set* of signaling systems then *P* ***and Π are operationally equivalent*** if *P* and ∪Π are operationally equivalent.

∪Π is the single system whose signals are exactly those carried by at least one signaling system in Π. For instance Π = {{*p*1}, {*p*2}} corresponds to the joint signaling capacities of two persons, one who only detects the presence of triangularity and one who only detects the presence of redness. (15) reflects the fact that that pair of signaling systems can, by its combined reports, tell us exactly as much about the state it is in as any that is operationally equivalent to {*p*1, *p*2}, for instance the signaling system {*p*1, *p*2, *p*4} as defined at 2.1(k).

(m) The final definition corresponds in this system to unity of consciousness i.e. gives a behaviouristic ersatz for it:

1. Let *P* be a signaling system with a non-empty subset *Y* of all its {*f*}-signals, a non-empty subset *Y*\* of all its {*f*\*}-signals, a subset *Z* of all its {*f*, *f*\*}-signals, and a subset *X* of all other signals. So *P* = *X* ∪ *Y* ∪ Y\* ∪ Z. Then if *P* is operationally equivalent to {X ∪ Y, X ∪ Y\*} we say *P* ***has distributed awareness of f and f\****. Otherwise *P* ***has unitary awareness of f and f\*.***

Imagine a patient of some sort of idealized commissurotomy who was never simultaneously aware of visible shape and tactile shape. Rather there were two ‘centres of consciousness’ that received information from the visual and the haptic channel respectively; but the two centres could co-operate on motor functions to produce a signal. Let him be asked to report on the look and feel of a range of small objects arranged haphazardly on a large table in front of him. We could then treat his reports on what looks square and what feels square as signals whose values supervene on the distributions of *f* = *looking square* and *f*\* = *feeling square* over these objects. But any report of his on the distribution of visual *and* tactile squareness would depend entirely on two types of separately reportable judgments, each coming ‘from’ just one or the other centre, the first type being concerned solely with what looks square and the second type being concerned solely with what feels square. (So he could not, for instance, possess the {*f*, *f*\*}-signal corresponding to the report that something both looks and feels square). And in this case he has distributed awareness of *f* and *f*\*.

Informally, the point of distributed awareness is that if a signaling system has distributed awareness of *f* and *f*\* then it makes no difference to its capacity to discriminate states that *one* signaling system is sensitive to the distribution of *f* *and* of *f*\* rather than a different signaling system’s (or a different hemisphere’s) being sensitive to each: having one such distributed witness to tell us whether something is red and whether something is triangular carries no informational advantage over dividing this labour between two witnesses, one for each property.

That completes the exposition of the model. Its elements are: (a) objects and their properties (b) states of the world (c) signals (d) {*f*}-signals (e) *F*-signals (f) signaling systems (g) object-insensitivity and object sensitivity (h) separation (j) P-indiscernibility (k) operational equivalence between signaling systems (l) operational equivalence between a signaling system and a set of signaling systems (m) distributed and unitary awareness. I now argue that unitary awareness entails object-sensitivity.

*2.2 Formal argument*

The argument has two premises: both are straightforward consequences of the foregoing.

1. ***Object-insensitivity implies separation***: For any object-insensitive signaling system *P*, if *q*, *q*\* ∈ *P* and *q* is a non-trivial {*f*}-signal and *q*\* is a non-trivial {*f*\*}-signal, then *P* separates *f* and *f*\*.[[5]](#footnote-3)
2. ***Separation implies distributed awareness***: If *P* has an {*f*}-signal and an {*f*\*}-signal and *P* separates *f* and *f*\* then *P* has distributed awareness of *f* and *f*\*.[[6]](#footnote-4)

A word about (18). It says that a signaling system *P* that separates *f* and *f*\* cannot distinguish states of the world any more finely than two signaling systems of which one has only *P*’s sensitivity to *f* and the other has only *P*’s sensitivity to *f*\*. As far as its discriminatory capacities are concerned, nothing is gained from its *f*-detecting capacities and its *f*\*-detecting capacities being combined into one agent as opposed to their being distributed over two.

Section 1 illustrates this point. If the objects are positions in space and the properties *f*1 or *f*2 apply to a position if something red or something triangular is located there respectively, then the horse is a signaling system with signals *p*1, *p*2 corresponding to ‘Red!’ and ‘Triangle!’ respectively. If responses in *L* exhaust its signals, then it separates *f*1 and *f*2 if and only if it lacks the signal *p*3 i.e. if and only if it *can’t* correctly answer the question ‘Red Triangle?’ (18) implies that if it can’t, it is as sensitive to the distribution of shapes and colours as it would be on the fragmentary hypothesis. Since *one* person who sees both shapes and colours *can* make more distinctions than that, the fragmentary hypothesis follows from the horse’s *in*capacity to answer ‘Red Triangle?’

The conclusion follows straightforwardly from (17) and (18).

1. ***Unitary awareness implies object-sensitivity***. If *P* is unitarily aware of any *f* and *f*\* then *P* is object-sensitive.

(19) is the advertised connection between unity and objectivity as applied to this model. If a signaling system discriminates amongst distributions of two properties in a way that two specialized observers (or one commisurotomized observer) could not, then it is sensitive to *which* object has either property or whether the *same* object has both. That is how a recognizable distortion of the main idea of Kant’s Deduction applies in this very bare and abstract setting.

*3. Comments*

But the differences are more interesting than the similarities and it is worth mentioning the most obviously distinctive features of (a) unity and (b) objectivity as explicated here.

(a) First, a signaling system can have more or less unitary awareness, whereas the unity that interested Kant (‘transcendental unity of apperception’) is either there or not. In one place, for instance, Kant explains it as demanding that:

The *I think* must *be able* to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much to say that the representation would be impossible for me or else at least would be nothing for me.[[7]](#footnote-5)

Either the ‘I think’ accompanies *all* your representations or it doesn’t. But distributed and unitary awareness are always relative to properties: the commisurotomized person, for instance, might be representable as a signaling system that has distributed awareness of *feeling square* and *looking square* but unitary awareness of *looking red* and *looking square*. So we can partially order signaling systems according to whether they are more or less unitary. It doesn’t make sense to call subjects ‘more’ or ‘less’ unified in Kant’s sense.

Second: as the passage also reveals, Kantian unity involves *awareness* of some other kind of unity. It is not enough that these representations *be* mine; I must *be able to judge* that they are. Unitary awareness makes no such demand because it isn’t mentalistic at all: whether a system has unitary awareness of these or those properties depends only on correlations between its outputs (signals) and its inputs (possible states of the world). What happens in between is immaterial. A signaling system might have unitary awareness without being conscious at all.

A third, more interesting difference is that modulo those first two, unitary awareness is more demanding than Kantian unity on one reading of the latter. Consider this famous example of William James’s, itself adapting one of Brentano’s:

Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them into a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence.[[8]](#footnote-6)

The analogy with unity of consciousness is supposed to be that in a disunified consciousness there is separate awareness of each word, but no awareness of the whole sentence. R. P. Wolff glosses this passage as follows:

The fact is that one consciousness of twelve words is not the same as twelve consciousnesses of one word each. Following Kant’s terminology, we may characterize the difference by saying that one consciousness of all twelve word binds them together, or conceives them as a unity.[[9]](#footnote-7)

On the Wolffian reading it is simply awareness of *all* of the words that constitutes unity. But that unity is consistent with distributed awareness in my sense. Let the sentence be ‘The correct occasions of its use are those attended by painful stimulation.’ Then a signaling system, for instance a text-scanning device, might have signals reflecting the presence of a token of each of the twelve types of word in the sentence. Supposing the objects in question to be regions of the printed page, and the properties to be *f*1 = *bears an instance of the word ‘The’*, *f*2 = *bears an instance of the word ‘correct’* … up to *f*12 *= bears an instance of the word ‘stimulation’*, then for 1 ≤ *i* ≤ 12, let the {*fi*}-signal *pi* take value 1 if and only if some object in the domain has the property *fi*.[[10]](#footnote-8) Finally, let *p* take value 1 if and only if *each* of the properties *f*1, … *f*12 is borne by some object i.e. [*p*] = ∩1 ≤ *i* ≤ 12 [*pi*]. Then the signaling system {*p*1, … *p*12} plainly has distributed awareness of *fi* and *fj* for any *i*, *j*. But {*p*, *p*1, … *p*12} – a scanner that can signals the presence of each word *and also* the presence of *all* of them – *also* has distributed awareness of *f*i and *fj* for any *i*, *j*. That corresponds to the case in which James’s twelve men could talk to one another – like the men in the horse – or to a thirteenth: at least one of them would be conscious *of all twelve words* and thus constitute unity in Wolff’s sense.

But consciousness *of all twelve words* is not consciousness *of the sentence* i.e. of the word *order*; and it is this second kind of unity that unitary awareness demands. For instance, a signaling system {*p*\*, *p*, *p1*, … *p1*2}, where the {*f*1, *f*2}-signal *p*\* takes value 1 if and only if the word ‘The’ is to the left of the word ‘correct’, has unitary awareness of *f*1 and *f*2.[[11]](#footnote-9) In terms of James’s example: even if the individuals could talk to one another, they might all have awareness of all of the words in the sentence; but still none would be aware of the sentence itself. It is awareness of the sentence, and not the merely aggregative awareness that constitutes Wolffian unity, that distributed awareness in my sense is supposed to be ruling out.[[12]](#footnote-10)

For a fourth contrast, now with Cartesian unity, note that unitary awareness is independent of the existence of a simple and/or substantial subject of all experience. Broadly speaking, this is because unitary awareness is behaviouristic. More specifically, a signaling system that is unitarily aware of all properties to which it is sensitive might possess a very complicated internal structure in which, as in Leibniz’s mill, no self might be found. Conversely, if we can imagine a simple unified subject at all, then we can imagine one that has distributed awareness of *all* of the properties that it registers. Suppose God can distinguish states in Table 1 with absolute precision because He has at least the four signals *p*11, *p*12, *p*21 and *p*22, defined as follows:

1. *p*11 (*s*) = 1iff *o*1 ∈ *f*1 (*s*)
2. *p*12 (*s*) = 1 iff *o*2 ∈ *f*1 (*s*)
3. *p*21 (*s*) = 1 iff *o*1 ∈ *f*2 (*s*)
4. *p*22 (*s*) = 1 iff *o*2 ∈ *f*2 (*s*)

The sixteen possible distributions of values of these signals correspond to the sixteen possible states; moreover *p*11 and *p*12 are both {*f*1}-signals and *p*21 and *p*22 are both {*f*2}-signals. It follows trivially from these facts and (13) that *any* signaling system *P* such that {*p*11, *p*12, *p*21, *p*22} ⊆ *P* separates *f*1 and *f*2 and so by (18) has distributed awareness of them: we could learn no more from *P* than from two signaling systems {*p*11, *p*12} and {*p*21, *p*22}. But clearly this is consistent with P’s simplicity and substantiality in so far as any sense attaches to those words at all.[[13]](#footnote-11) Hence also we should distinguish the kind of unity that I am saying is sufficient for object-sensitivity from the kind of unity that others have taken to be necessary for it, as in the well-known Buddhist argument that the self is a substance because the same object may be seen and touched.[[14]](#footnote-12)

(b) Let me also mention two contrasts between object-sensitivity and related notions. First, object-sensitivity need not involve a signaling system in any commerce with objects that are *external* to it, *either* in the sense of being spatially exterior *or* in the sense of existing or possibly existing unobserved. It is consistent with (19) that the objects to which a unitarily aware signaling system is sensitive are sense data or Berkeleian ideas, if these are considered as distinct entities that bear properties[[15]](#footnote-13). So there is no prospect of establishing our talk of external objects on any firmer footing than that reached by the end of Hume’s skeptical discussion[[16]](#footnote-14).

But to see the argument as anti-skeptical is to miss its point. It is not trying to convince anyone that there *are* external objects. Rather it takes a ‘side-on view’ of the signaling system or ‘subject’ that *takes it for granted* that there are objects, external or otherwise, isolates senses in which an abstract signaling system may be (i) unified and (ii) objective, and then connects (i) and (ii). In this respect the approach is more Strawsonian than Kantian.[[17]](#footnote-15)

The second thing worth distinguishing from object-sensitivity is any *semantic* relation between signals and objects. Calling a signaling system object-sensitive is to not saying that its signals denote objects. One *might* say that; but nothing in the content of (19) or the argument for it depends on that. On the other hand, (19) does have interesting *application* to a primitive language for which the line between objective-insensitivity and object-insensitivity coincides with an important semantic distinction. I turn finally to this.

*4. Idealistic languages and the self*

Imagine (again) a community of speakers whose language *L*1 consists of a stock of expressions similar to those in fragment (1)-(6) of *L*: a range of simple expressions, learnt by children through ostensive conditioning[[18]](#footnote-16), to which the mature speaker is disposed to assent in the sensible presence of something with one or other of a corresponding range of properties; and otherwise to dissent. Signals of assent and dissent aside, *L*1 thus consists entirely of simple ‘observation sentences’: one-word expressions assent or dissent to which depends entirely on the speaker’s contemporary surroundings.[[19]](#footnote-17)

Speakers of *L*1 are object-insensitive signaling systems. For each sentence of the language, the disposition to assent to it or dissent from it amounts to the possession of a signal taking value 1 if and only if *something* in the locality has this or that property – never mind which ones, or how many.

Furthermore, nothing in the linguistic practices of *L*1-speakers gives us any reason to think of its words as predicates with divided reference over a *range* of objects. Thus assent to ‘Triangle?’, say, need not be interpreted as asserting that something triangular is around. Of course it *could* be, since this would get its assertion-conditions right: the speaker assents to ‘Triangle?’ if and only if something triangular *is* around. But we could equally interpret them as predicates that apply periodically to *themselves*. That is, we could interpret assent to ‘Triangle?’ as asserting ‘*I* am in a triangle-detecting state’, since this too gets the assertion-conditions right. On this second interpretation the speakers are solipsists. The point is not that they *are* solipsists, nor that we are *forced* to interpret them as such, but that we could. If a language is thus interpretable i.e. as describing only modes of consciousness, I’ll call it *idealistic*.

Since *L*1-speakers are object-insensitive, they have distributed awareness; the use of *L*1 to distinguish states of the world is independent of any unification in the sensory capacities that underlie it. Whether, for instance, the visual input that issues in ‘White!’ and the haptic one that issues in ‘Furry!’ operate on the speaker’s linguistic output through entirely independent channels is irrelevant to these discriminatory capacities, since any speaker of *L*1 is operationally equivalent to one in which they do.

Consider now a language *L*2 that shares its vocabulary with *L*1 but admits of compound sentences, i.e. ones concatenating two simple expressions (‘Red triangle?’ etc.), to which the speaker assents if and only if something in the locality has both properties corresponding to the constituents. (Thus e.g. the speaker assents to ‘Red triangle?’ if and only if some (*one*) such thing is both red *and* triangular.) Children are taught all simple and a few compound expressions of *L*2 through ostensive conditioning, as with *L*1, but can then *spontaneously* grasp novel compounds. For instance, a child who learns ‘Red?’, ‘Triangular?’ and ‘Blue square?’ then grasps ‘Red square?’ without prior exposure to *it*.

Speakers of *L*2 are object-*sensitive*. Assent to some such compound expression (corresponding e.g. to *p*3 at 2.1(e)) does not supervene on whether *something or other* has one, and *something or other* has the other, of the properties corresponding to its constituents (e.g. corresponding to *p*1 and *p*2 at 2.1(c) and 2.1(d)).

Furthermore, the step from *L*1 to *L*2 creates a difficulty for the solipsistic semantics. We could not interpret ‘Red Triangle!’ as conjoining the solipsistic interpretations of ‘Red!’ and ‘Square!’ in *L*1, since the last two sentences are both assented to, but the first is dissented from, when the speaker faces a red square and a blue triangle. Of course we *could* interpret ‘Red Triangular!’ as asserting that the speaker is ‘red-triangle-detecting’, where this describes some *third* state that he is in just when confronted with a red triangle. But that frustrates the explanatory ambitions that we might reasonably expect *interpreters* to treat as desiderative[[20]](#footnote-18): we cannot now account for the learner’s grasp of novel compounds with familiar elements. ‘Red’ denotes the speaker when he is in one state; ‘triangle’ when he is in a second; but how can knowing these two things by themselves tell any learner what *third* state he has to be in for ‘red triangle’ to denote him? And yet this is something on which speakers of *L*2 are naturally able to pick up.

An interpretation based on external objects faces no such difficulties. One natural such approach interprets ‘red’ and ‘square’ as predicates with divided reference (extension) over various objects, so that assent to ‘Red?’ says that some one object (at least) is red. Then we can understand ‘red square’ as a complex predicate true of exactly those objects to which both constituent predicates apply. This recovers the assertion-and assent-conditions of compounds. And it explains learners’ capacity to grasp new compounds having learnt at least one old one: to understand ‘Red square’, for instance, it suffices to grasp ‘red’ and ‘square’ and to apply a general rule that yields the denotation of any compound given those of its constituents, namely that *o* belongs to the extension of ⎡*FG*⎤ iff *o* belongs to the extension of ⎡*F*⎤ *and* to the extension of ⎡*G*⎤; and this is something that one could pick up from having learnt *other* compounds. So in passing from *L*1 to *L*2 we cross the line from an idealistic language to a more object-involving one.[[21]](#footnote-19)

And we have also crossed the line from distributed awareness to unitary awareness. No speaker of *L*2 is operationally equivalent to a collection of speakers over whom, or to one speaker over whose cerebral hemispheres, its colour- and shape-detecting capacities are distributed. If we imagine *L*1 and *L*2 as successive stages in the development of a single language, this means that the transition from that earlier idealistic stage to this later non-idealistic one also marks, not the point at which it happens, but the point at which it matters, that its speakers have unity of consciousness.

*5. Conclusion*

This paper attempted to construct a very simple and highly idealized model for which a relatively precise and non-psychological explication of the Kantian thesis holds true. Section 4 illustrated its consequences for an imaginary language; but that is very far from claiming that the model can be refined so as to tell us anything worth knowing about real people. Many such refinements would be necessary to show that, starting with: (i) replacing supervenience of signals on states with some weaker, perhaps probabilistic relation; (ii) allowing signals to take more than two values, to cope with expressions of uncertainty; (iii) relatedly, treating separately of borderline cases, where one assents to a disjunction without assenting to either disjunct; (iv) allowing objects to exist at only some states of the world; (v) accounting for ‘Gestalt’ perception, e.g. seeing that something is red *and* square without seeing either that it is red or that is square.[[22]](#footnote-20) None of that is possible in the space that I have here. Still, I think it does already tell us something about the robustness and generality of Kant’s insight that it is both recognizable and true even in as austere and rudimentary a model as that which I *have* been able to describe.

**References**

Anscombe, G. E. M. 2011. Plato, soul and ‘the unity of apperception’. In Geach, M. and L. Gormally (ed.), *From Plato to Wittgenstein: Essays by G. E. M. Anscombe*. Exeter: Imprint Academic: 25-33.

Berkeley, G. 1996a [1734]. *Principles of Human Knowledge*. Second edn. Reprinted in M. Ayers (ed.). *George Berkeley: Philosophical Works*. London: Everyman: 71-153.

———. 1996b [1734]. *Three Dialogues between Hylas and Philonous*. Third edn. Reprinted in M. Ayers (ed.). *George Berkeley: Philosophical Works*. London: Everyman: 155-252.

Burnyeat, M. F. 1976. Plato on the grammar of perceiving. *Classical Quarterly* NS 26: 29-51. Reprinted in his *Explorations in Ancient and Modern Philosophy*. Cambridge: CUP 2012: 70-98.

Evans, G. 1975. Identity and Predication. *Journal of Philosophy* 72: 343-63. Reprinted in his Collected Papers. Oxford: OUP: 25-48.

Field, H. 1974. Quine and the correspondence theory. *Philosophical Review* 28: 200-28.

Ganeri, J. 2000. Cross-modality and the self. *Philosophy and Phenomenological Research* 61: 639-57.

Hume, D. 1949 [1740]. *Treatise of Human Nature*. Edited with an analytical index by L. A. Selby-Bigge. Oxford: Clarendon Press.

James, W. 1890. *Principles of Psychology*. 2 vols. New York: Henry Holt.

Kant, I. 2000 [1782/7]. *Critique of Pure Reason*. Ed. and tr. P. Guyer and A. W. Wood. Cambridge: CUP.

Kemp Smith, N. 1923. *A Commentary on Kant’s ‘Critique of Pure Reason’*. London: MacMillan.

Pereboom, D. 2006. Kant’s metaphysical and transcendental deductions. In Bird, G. (ed.), *A Companion to Kant*. Oxford: Blackwell: 154-68.

Quine, W.V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

———. 1970. Grades of theoreticity. In Foster, L., and J. W. Swanson (ed.), *Experience and Theory*. London: Duckworth: 1-17.

———. 1974. *The Roots of Reference*. Chicago and La Salle, Ill.: Open Court.

———. 1981. Empirical content. In his *Theories and Things*. Cambridge, Mass. Harvard University Press: 24-30.

Strawson, P. F. 1959. *Individuals*. London: Methuen.

———. 1966. *The Bounds of Sense*. London: Routledge.

Sullivan, P. 2004. Frege’s logic. In Gabbay, D. M. and J. Woods (ed.), *Handbook of the History of Logic vol. 3*. Elsevier: 659-750.

Wittgenstein, L. 1978. *Remarks on the Foundations of Mathematics*. Third ed. Ed. G. H. von Wright, R. Rhees and G. E. M. Anscombe. Tr. G. E. M. Anscombe. Oxford: Blackwell.

Wolff, R. P. 1967. A reconstruction of the argument of the Subjective Deduction. In Wolff, R. P. (ed.), *Kant: A Collection of Critical Essays*. New York: Anchor: 88-133.

1. Cf. Plato, *Theaetetus* 184d; Burnyeat 2012: 96-7, 97n. 61; Anscombe 2011. [↑](#footnote-ref--1)
2. Nothing essential hangs on the restriction to properties (i.e. to the exclusion of relations). Everything could be stated in terms of relations too, but this would complicate the exposition, since the extensions would now have to be sets of ordered n-tuples and not just sets of objects. See further n. 4 below. [↑](#footnote-ref-0)
3. Evans’s term (Evans 1975) of course comes from Strawson (1959: 202). But Strawson seems to me to mean slightly more by it: for him, feature-*placing* expressions (e.g. ‘Cold!’ and ‘It’s raining!’ but also ‘Red there!’ and ‘Warm here!’) are slightly more sophisticated than the signals of a feature detector in that the former may *place* features within the local environment whereas the latter always just register their presence in it. I should add that object-insensitivity as stated in (12) is slightly more demanding than Evans’s feature-placing in another way: a feature-placing language might contain syntactically simple expressions corresponding one to the presence of something white, one to the presence of a rabbit, *and* a third corresponding to the presence of a white rabbit. Object-insensitivity rules this out: if *f* is the property of being a local rabbit and *f*\* is the property of being local and white, then *any* {*f*, *f*\*}-signal, ‘simple’ or not, depends only on whether something local is white and something local is a rabbit: so if object-insensitivity holds then no signal can take value 1 exactly at states in which the some one local thing is a white rabbit. [↑](#footnote-ref-1)
4. The restriction to monadic properties thus illustrates a mode of involvement with objects that doesn’t require ‘relational’ signals. {*p*1, *p*2, *p*3} – these signals as defined at 2.1(c)-(e) – is object-sensitive but only signals the distribution of monadic properties. Some writers take the view that what Kant regarded as logic – propositional and monadic predicate logic – has no need of objects, on the grounds that *nested* quantification – and hence variable letters, traditionally the mark of objectual involvement – is only necessary in a relational language, since any monadic formula is equivalent to one without nesting (Sullivan 2004: 705). The example shows that even *within* the realm of propositional and monadic predicate logic there may be grades of objectual involvement. [↑](#footnote-ref-2)
5. Proof: Let *p* ∈ *P* be an {*f*, *f*\*}-signal and let *s* and *s*\* be such that *p* (*s*) ≠ *p* (*s*\*) but *q* (*s*) = *q* (*s*\*) and *q*\* (*s*) = *q*\* (*s*\*). It follows from (12) and *p* (*s*) ≠ *p* (*s*\*) that either (i) ¬(*f* (*s*) = Ø ↔ *f* (*s*\*) = Ø) or (ii) ¬(*f*\* (*s*) = Ø ↔ *f*\* (*s*\*) = Ø). Suppose (i) is true. Then for an arbitrary state *t*, either (iii) (*f* (*t*) = Ø ↔ *f* (*s*) = Ø) or (iv) (*f* (*t*) = Ø ↔ *f* (*s*\*) = Ø). If (iii) is true then by (12) *q* (*t*) = *q* (*s*); if (iv) i*s* true then by (12) *q* (*t*) = *q* (*s*\*), hence again *q* (*t*) = *q* (*s*). So *q* (*t*) = *q* (*s*) for any *t*, contradicting the non-triviality of *q*. So (i) is false; by parallel reasoning from the non-triviality of *q*\*, (ii) is false. Hence the initial supposition was false; so for any {*f*, *f*\*}-signal *p* ∈ *P*, there is *Q* = {*q*, *q*\*} such that *q* is an {*f*}-signal and *q*\* is an {*f*\*}-signal, *Q* ⊆ *P*, and ∀*s*∀*s*\* (∀*r* ∈ *Q* (*r* (*s*) = *r* (*s*\*)) → *p* (*s*) = *p* (*s*\*)): so by (13), *P* separates *f* and *f*\*. [↑](#footnote-ref-3)
6. Proof: suppose *P* separates *f* and *f\** but has a signal *z* that gives different values in two states *s* and *s*\* over which all signals in *X* ∪ *Y* ∪ *Y\** agree. Then *z* ∈ *Z* since *P* = *X* ∪ *Y* ∪ *Y\** ∪ *Z*. So z is an {*f*, *f*\*}-signal. But since *P* separates *f* and *f\**, it follows from (13) that *z* (*s*) = *z* (*s*\*). Therefore if all signals in X ∪ *Y* ∪ *Y\** agree over two arbitrary states *s* and *s*\* then all signals in *P* agree over them. Conversely, it obviously follows from *X* ∪ *Y* ∪ *Y\** ⊆ *P* that if all signals in *P* agree over two arbitrary states *s* and *s*\* then so do all signals in *X* ∪ *Y* ∪ *Y\**. It follows from (14) that that two states are *P*-indiscernible if and only if they are (*X* ∪ *Y* ∪ *Y\**)-indiscernible, from 2.1(k) that *P* and *X* ∪ *Y* ∪ *Y\** are operationally equivalent, from (15) that *P* and {*X* ∪ *Y*, *X* ∪ *Y\**} are operationally equivalent and from (16) that P has distributed awareness of *f* and *f*\*. [↑](#footnote-ref-4)
7. Kant 2000 [1787]: B131-2; emphasis in original. [↑](#footnote-ref-5)
8. James 1890 vol. I: 160. Kant himself uses a similar example in a slightly different connection (Kant 2000 [1782]: A352). [↑](#footnote-ref-6)
9. Wolff 1967: 91. [↑](#footnote-ref-7)
10. For this example to work, we need to relax the assumption made at 2.1(b) that no property excludes any other: in this case, nothing can be both *fi* and *fj* if *i* ≠ *j*. But the example illustrates how that simplifying assumption was unnecessary anyway: nothing in the proofs of (17) or (18) relied on the assumption that *fi* and *fj* were compatible. [↑](#footnote-ref-8)
11. Let the visible regions *o*1, … *on* be ordered such that *oi* is to the left of *oj* iff *i* < *j*. Then the signal *p\** takes value 1 at a state s if and only if *oi* ∈ *f*1 (s) and o*j* ∈ *f*2 (s) for some *oi*, *oj* s.t. *i* < *j*; plainly this is an {*f*1, *f*2} property. But there are states of the world *s* and *s*\* such that *p*1 (*s*) = *p*1 (*s*\*), *p*2 (*s*) = *p*2 (*s*\*) and *p* (*s*) = *p* (*s*\*), but *p\** (*s*) ≠ *p\** (*s*\*): e.g. let *s* be the state in which the sentence reads ‘The correct occasions of its use are those attended by painful stimulation’ and let *s*\* be that in which it reads ‘Correct occasions of its use are those attended by the painful stimulation’. So {*p\**, *p*, *p1*… *p1*2} is not operationally equivalent to {{*p1*, *p*3, … *p1*2}, {*p2*, *p*3, … *p1*2}} i.e. the former has unitary awareness of *f*1 and *f*2. [↑](#footnote-ref-9)
12. The contrast drawn here between unitary awareness and Wolffian unity also distinguishes unitary awareness from an *inferential* vision of unity. Pereboom for instance suggests that ‘the unity consists in certain intimate ways in which representations in a subject are typically related. Perhaps the key aspect of this unity is that a single subject’s representations are inferentially integrated to a high degree, by contrast with representations across distinct subjects’ (2006: 159). One might take this to mean that one draws conclusions from several representations that do not follow from any one: I learn by sight (A) that somebody is wearing a hat (∃*xf*1*x*) and by hearing (B) that somebody is shouting (∃*xf*2*x*); I infer something that neither sight nor hearing could have taught me by itself i.e. (C) that somebody is wearing a hat *and* somebody is shouting (∃*x*1∃*x*2 (*f*1*x*1 ∧ *f2x*2). But a device with signals *p*A, *p*B, *p*C corresponding to these three propositions might still have distributed awareness of *f*1 and *f*2 for the same reasons as apply to Wolff. [↑](#footnote-ref-10)
13. This example also illustrates a signaling system that is object-sensitive as well as separated; so the converse of (19) is false. This marks another contrast with Kant, at least on the reading that makes his notions of unity and objectivity *mutually* dependent (as in Kemp Smith 1923: 253). [↑](#footnote-ref-11)
14. Ganeri 2000. As well as its proceeding in the opposite direction from mine, a further difference between the Buddhist line and mine is that the former involves *cross*-modal capacities i.e. recognition via e.g. sight and touch, whereas the latter is equally concerned with *intra*-modal distributed awareness, as in the Trojan horse. [↑](#footnote-ref-12)
15. Whether Berkeley has the right to say that is another question. Berkeley disagrees with Hume (1949 [1740]: I.iv.6) in accepting a simple unitary subject of experience (1996a [1734]: sect. 27) and in claiming that ideas can only exist ‘in’ such a subject (*ibid*. sect. 33). Perhaps the only way to make clear sense of this is to treat the having of an idea as a state or property of the self, even though this is something that he expressly denies (1996b [1734]: p. 227). In that case Berkeley is committed to saying that the language of perceptual experience is *idealistic* in the sense of section 4 below; in that case (19) *does* apply. [↑](#footnote-ref-13)
16. Hume 1949 [1740]: I.iv.2. [↑](#footnote-ref-14)
17. Strawson 1966: 31-2. Of course it also follows Strawson in largely ignoring transcendental (or empirical) psychology. [↑](#footnote-ref-15)
18. As described in Quine 1974: 41-5. [↑](#footnote-ref-16)
19. This is in roughly Quine’s 1981 sense (1981: 25): ‘roughly’ because observationality in that paper (i) comes in degrees and (ii) is relative to neural stimulations not *distal* stimuli. Modulo these differences, the criterion of observationality in the text agrees with this one of Quine’s in the sense of being individualistic. Contrast the more communitarian conception at Quine 1960: 43. [↑](#footnote-ref-17)
20. Interpreters, that is, as opposed to *translators*. See Evans 1975: 25-7. [↑](#footnote-ref-18)
21. Evans argues that compound formations alone do not motivate treating ‘red’ and ‘square’ as predicates, because nothing settles what it is for such a predicate *not* to be true of something – the only states where a speaker dissents e.g. to ‘Red?’ are those where *nothing* is red, and this is not necessary for there to be *something* that is not red (1975: 31-3). But even if we need not understand ‘red’ as a predicate with some *particular* divided reference, it remains true that any semantic account of it must *somehow* mention distinct objects, so at least in this sense the compound language is object-involving in a way that the simple one is not. For instance, when we are dealing with equivalents in *L*2 of our mass terms (‘red’, ‘wine’), we *might* treat them each as denoting a *distinct* scattered individual, and a compound of two as denoting the common part of the denotations of its elements (as Quine (1970: 8-9) suggests for our actual mass terms). Or we might treat ‘red’ and ‘square’ as *partially* denoting various sets of objects (Field 1974: 209f.); but again, if we wish to account for their interaction in compound constructions, these must all be sets with more than one element. [↑](#footnote-ref-19)
22. For (v) see Wittgenstein 1978: VII-63-5, 68-9. Wittgenstein writes: ‘I can imagine someone saying that he saw a red and yellow star but did not see anything yellow–because he sees the star as, so to speak, a conjunction of coloured parts, which he cannot separate’ (1978: VII-65). Such a person might then have an {*f*1, *f*2}-signal (‘Yellow-and-red!’) without having *any* {*f*1}- or {*f*2}-signals, and so would count as having distributed awareness of *f*1 and *f*2 and hence being object-sensitive; but this is very unintuitive. One response might be to relativize the individuation of properties to signaling devices in such a way that if *P* has any {*F*}-signals then *P* has an {*f*}-signal for each *f* ∈ *F*. [↑](#footnote-ref-20)