



UNIVERSITY OF ILLINOIS PRESS

North American Philosophical Publications

On the Identity of Sensations and Physiological Occurrences

Author(s): Richard Routley and Valerie Macrae

Source: *American Philosophical Quarterly*, Vol. 3, No. 2 (Apr., 1966), pp. 87-110

Published by: University of Illinois Press on behalf of the North American Philosophical Publications

Stable URL: <http://www.jstor.org/stable/20009197>

Accessed: 19-06-2016 20:26 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



North American Philosophical Publications, *University of Illinois Press* are collaborating with JSTOR to digitize, preserve and extend access to *American Philosophical Quarterly*

I. ON THE IDENTITY OF SENSATIONS AND PHYSIOLOGICAL OCCURRENCES

RICHARD ROUTLEY AND VALERIE MACRAE

I

THE main hypothesis to be considered is, in initial formulation:

(A) *Sensations are physiological occurrences*
“Physiological” is used in its normal sense, so that physiological occurrences include both bodily occurrences like the damaging of tissues and neurophysiological occurrences. The frequently discussed formulation “Sensations are brain processes” over-restricts the class of physiological processes relevant to sensations.¹ When (A) is asserted it is generally intended as an identity. To make the identity explicit (A) needs reformulation; for at face value it expresses an *inclusion*. Identity is a symmetrical relation; but the truth-value of “Physiological occurrences are sensations” may well differ from that of (A). Other indeterminacies in (A) appear as soon as we start asking about equivalent, or synonymous, reformulations. Improved formulations which avoid some of these problems are:

(B₁) *Sensations of sort h(g) that belong to or are present in x are identical with physiological occurrences of sort g in x, for any x and any g.*

(C₁) *x has a sensation of sort h(g) iff a physiological occurrence of sort g takes place in x, for any x and any g.*

What, however, are the ranges of the variables? Consider the variable ‘x’; an obvious choice is to let this variable serve as placeholder for referring expressions which refer to particular animals, including humans. The predicate-variable ‘g’ serves as placeholder for certain descriptions (most details of which are not known) of physiological processes; ‘h’ stands for a function which correlates described physiological occurrences with sensa-

tions; and so ‘h(g)’ ranges over descriptions of sensations. Actually, the relation between a subclass of physiological occurrences and sensations may not be so simple that it can be expressed by a single-valued function, or even by an explicit function like *h*. (C₁) has the *form* of a standard type of reduction sentence; whether it is or not depends on the tricky issue as to whether it contains the requisite observational and theoretical expressions (see Sect. II).

Before discussing the relevant senses of “iff” and “are identical with,” it is important to clarify the sense and designation of “sensations” and to distinguish *sorts* of sensations. A quick glance at the etymology of “sensation” (*vide* the OED) shows that the word has been used in quite a wide way since about 1500 to cover items like feelings, e.g., of pride and elation, and suppositious items like visual, tactual, kinaesthetic sense-impressions or -data, as well as physically localizable, often disturbing, person-limited bodily sensations. To avoid confusion and the necessity of making qualifications “sensation” will be used, from now on, in the sense in which it is restricted to refer to (occurrences of) items of this last cluster, that is, to sensations proper. Included in sensations proper are: bodily pains of all sorts; itches, ticklings, shivers, numbness, and blackout sensations; internal pressure, orgasm, giddiness, thirst, nausea, seasickness, suffocation, dazzle, and after-image sensations. The distinguishing marks of sensations proper explained below are vital because:

(i) These marks serve to delimit the cluster and reveal important similarities between various different sorts of sensations, e.g., between sensations as different as bodily pains, giddiness, and the having of after-images.

(ii) For each group of features there are many mental items, such as emotions, thoughts, volitions, and mental dispositions, which differ from

¹ Similarly, central state materialisms soon have to be widened to include other bodily processes and behavior, e.g., that of endocrine systems which seem important in connection with emotions. Once the extension is made some of the barriers separating materialisms and analytical behaviorism are down.

sensations in respect to these features. Some of the features also serve to isolate sensations from perceptual observations.

(iii) These features make identity hypotheses look much more plausible and more testable when restricted to sensations than when extended to cover all mental items. The restricted hypotheses are more amenable to testing, and (as it already seems) we can establish an accurate temporal correspondence between and also similarities of patterns of sensations and physiological occurrences.

Restricted identity-hypotheses are more plausible for similar reasons (developed in setting out the marks of sensations). This is so for three reasons: (1) because of the precision with which the duration of sensations can be marked out; (2) because of their bodily locatability; and (3) because of the many ways in which they can be assimilated to physiological occurrences. For these reasons, restricted hypotheses like instantiations of (B_1), are not to be confused with full materialist hypotheses or accounts of mind, which endeavor to make a reduction—of a fairly uniform character and satisfying some or other (though often it is not clear which) of a range of possible reduction relations—of *all* mental items to physical or material data of *some* sort. Restricted hypotheses are *compatible* with more sophisticated many-strand theories different from materialism, indeed with analytical behaviorism which is not committed to offering a behavioristic account of sensations; though vindication of (A) would make serious inroads into dualisms.

So there are two suppositions we put up:

(a) Materialisms, especially central-state theories, are more plausible for sensations, just because of some special features of sensations, than for most (other) mental items; if it fails here then it fails entirely. Moreover, if it runs into difficulties over sensations then it is much more likely to run into difficulties, some of them transferable from the sensation case, over emotions, thoughts, and decisions.

(b) Sensations proper are *not* correctly classified as *mental* occurrences. Admittedly it is fairly standard nowadays to classify sensations, though not reflex actions, as mental occurrences; but the classification does not have a good pedigree (cf. Greek classifications and Cartesian difficulties over animals), and when in the past sensations did get in they tended to have second-class status. It is possible, by taking into account different senses of

“mind” and of “mental” and features of sensations, to develop something of a case for their reclassification. For instance, sensations don’t have much more relevance to a person’s mental, intellectual, or emotional states than might some of his physical characteristics, e.g., having lost a limb. A person who had no sensations could still have a mind. Also the question “Do rats have minds?” is a conflict-question but the question “Do rats have sensations?” is not. Consider, too, details introduced in discussing someone’s mental qualities, whether he has a good mind, what his mental state is. The conventional way of settling this conflict issue over whether sensations are mental occurrences or not, seems to draw much of its strength from dubious stream-of-consciousness or sensationist views on what it is to have a mind. If (b) is defensible, sensations, and relatively incorrigible occurrences, are very far from being paradigms of mental occurrences; and establishing of (B_1) would go little distance toward establishing a full materialist account of *mind*, though, if (a) is correct, refuting (B_1) would go a good way toward refuting it.

Important distinguishing marks of sensations, on which fuller defense of the restricted identity claim relies, are then:

(1) Sensations are items that we *have* at given times or over periods of time. They last fairly definitely specifiable periods of time. In general, one can say, with a high degree of precision, of a sensation that it’s just started or that it’s now gone. Sensations are *occurrences or processes*. They are not events in the narrow sense of “event,” neither are they like achievements which have a time of happening but no duration. Sensations are occurrences or perhaps, if they are sufficiently frequent or dominant, “states.” They are episodic and not dispositional though a person may be disposed to or liable to get certain (sorts of) sensations. Because sensations are occurrences, it is the *having* of after-images and not after-images themselves, i.e., processes and not the products, that count as sensations.

(2) All sensations either have a *bodily location*, which may, as in tingling or fever sensations, be quite extensive, or else are intimately connected, like after-imaging, with a bodily condition. To guarantee this point in the case of pains we take over the traditional distinction between mental pains, like anguish, which are better classed with emotions, and bodily pains. This distinction can be sharpened by the adoption of physiological criteria for real bodily pains.

(3) They happen to us; they are not something we do. In this respect especially, sensations may be likened to reflex actions, with which they are often connected, as for instance withdrawal from a cause of pain and avoidance of a source of dazzle. True, with all of them there are recipes or procedures for producing them: but having carried out these procedures and set up appropriate conditions, there is no stopping having them, except that in some cases there are counter-procedures we can use to prevent them provided the procedures are initiated early enough. We can have sensations accidentally or without intending to and when we don't want to have them; and they aren't produced just by *deciding* to have them without setting up the appropriate physical conditions connected with their occurrence. Consequently they are not put down to our agency, we are not held *responsible* for having them, though we may be held responsible either for causing or setting up the conditions for them either in ourselves or, more often, in others. Thus even the having of ordinary sensations is unconnected (except in contingent ways, e.g., in psychosomatic disorders) with personality and intellectual traits.

(4) Although some people are better than others at detecting, locating, and describing their own, and even others' sensations and sometimes at producing them, the having of sensations is not at present an exercise of skill or care or a matter for training; nor in general are their occurrences a part of or the result of discrimination, detection, or the gaining of information. Thus there is little scope for mistake, little liability to correction with regard to the having of sensations. These features set the having of sensations apart from perceptual observation, from seeing, watching, tasting, etc., where training is relevant, where skill can be gained and may be needed, where mistakes are common enough, some illusions are normal, and achievements occur. Through visual, tactual, and other processes we often gain information about external items and their relations, and sometimes about internal items; but the having of sensations does not provide us directly with such information though, since certain sensation patterns are regularly produced under certain conditions; we can sometimes get data about external conditions in this way. Despite these differences the precise boundary between sensations and sense perception may remain problematic if it is not noticed that kinaesthetic receptors serve a dual purpose (thus heat receptors appear both in physiological

accounts of shivering and fever sensations and in the causal-physiological story of the way we detect the temperatures of external items). It should also be noticed that much of our vocabulary for describing sensations is transferred from that used to specify observable processes or features (consider descriptions of pains and after-imaging). With sensations we do not do better than achieve comparisons with external items, comparisons which already require perceptual observation of such items. Sensations do not in fact have external correlates; no after-images on walls, no pains in red-hot elements.

(5) First-person sensation reports are *relatively* incorrigible; havers are not likely to be corrected. Mostly sensations are readily identifiable to the haver. Very many of them are describable in detail as to their sort, pattern, history, changes, intensity, duration, and location or apparent location. Indeed, a person who has sensations cannot help noticing that he has them if he has them with any intensity and provided there are no sufficiently important distractions. The qualifications exclude rare exceptions such as the wounded soldier or the person who has undergone lobotomy.

(6) When the appropriate conditions—conditions which vary with the sort of sensation but are standard and specifiable for most sorts of sensations—are fulfilled everyone who is in a normal physiological condition in the relevant respects has the sensation. After-imaging and angina pains provide typical examples.

(7) If asked to explain the occurrence of specific sensations we usually resort to *physiological-type-explanation* and indicate some physical source and cause. It is worthwhile separating out several explanations of this type:

(a) Explanations referring to extra-bodily agents or objects which explain the *onset* of sensations but not, as a rule, their continuation; e.g., Jack hit you with a shovel, and now you are seeing spots of light. (In certain cases, with rapidly recovering tissues, etc., the continuation of a sensation may be explained by continual contact with the external source.)

(b) Explanations referring to bodily non-cerebral regions and processes in or features of these regions, as when the tissues in your shoulder are damaged (through the impact of the shovel), cones of your eyes are bleached (through staring at the light). When combined with laws of such forms as, "tissues damaged (fatigued, . . .) in respect *R* take under conditions *C* time *t* to recover, exhibit pat-

tern P in recovering," facts under (b) can be used to explain the continuation of sensations, their duration, their approximate intensity, and their pattern. Consider after-imaging: given details of the structure of the eye, in particular the photochemistry of cones and rods and of the optical nerve; given recognized physical and chemical generalizations, and given in each case relevant initial conditions like properties of the item observed, its relation to the observer, and period of observation, one can give a fair explanation of, and can predict, after-imaging and features of it. One can predict the duration and appearance, size, shape, brightness, and saturation of the images had. To complete our explanations we turn to:

(c) Explanations referring to neurophysiological occurrences which tell a story, generally starting with reactions of nerve fibers in regions referred to in (b) and ending with processes occurring in the brain, and in particular in the relevant sensation areas, e.g., pain-centers or c-fibers. In the case of psychosomatic sensations (a) and (b) may not be relevant; though diagnosis of stresses on the patient may be.

A complete context-independent explanation of specific sorts of sensations, which should in the end explain most of their features, combines (a), (b), and (c). There seems no upper limit short of explanation of *all correctly* observed and reported features of sensations as our explanations progressively improve with increased physiological data and assimilation of it under laws. These explanations would at least discern regular correlations between given sorts of sensations and physiological occurrences and result in control over sensations. By applying the explanations sensations could be induced, prolonged, shortened, or alleviated. But when observations of relations between correlated processes (strictly observations of relation-instances) are precluded or seriously limited (as here), regular correlations, as linked with explanations and connected under time-directed laws, together with related manipulative control, constitute first-rate grounds for the presence of causal relations. Therefore, given all these, the physiological explanations would be what they seem to be—causal explanations. It is extremely difficult, perhaps impossible, to obtain general explanations or characterizations of the causes of sensations without introducing

physiological matters. Asked to explain² what caused a particular after-image we may say "Looking at the one light over there too long"; but asked to explain after-imaging in general it is not usually satisfactory to say "After-imaging is caused by looking at bright objects too long." Physiological explanation is needed.

II

Consider hypotheses (B_1) and (C_1). Abbreviating "sensations of sort $h(g_0)$ in x_0 " by "SENGHS" and "physiological occurrences of sort g_0 in x_0 " by "PROGS," we may write an instantiation of (B_1):

- (B) SENGHS are identical with PROGS; or
 $\hat{z}: (\text{SENGH } (z)) = \hat{z}: (\text{PROG } (z))$
- (B) states a contingent class identity rather like
- (D) Humans are featherless plantigrade bipeds; or
- (D_1) Genes are (large parts of) DNA molecules; and
not like
- (D_2) Brothers are male siblings

where (D_2) yields an analytic statement. Unless the identity in (B) is formulated more explicitly problems arise as to what statement (B) is intended to yield. How can (B) be made more explicit? A recent move is to invoke a sense/reference distinction; to distinguish

(E_1) "SENGHS" has the same sense as "PROGS,"
 from

(E_2) "SENGHS" has the same referents as "PROGS"

and to equate (B) with (E_2). (E_2), despite its merits as a reformulation, has some drawbacks: first, it isn't obvious how extensive the referent-class is; second, "same" also appears in (E_2). In fact the intended referent-class, e.g., of "SENGHS" is the class of all referents which exist at some time, i.e., the total extension; and this last class is *independent* of the sense of "SENGHS" except that the class is a subclass of the possible extension which is logically determined by the sense of "SENGHS." Consequently we can *consistently adjoin* to a specification of the total extension of a referring expression (e.g., the class of all sometime actual featherless bipeds or "humans") a specification of the sense of the

² Here "explanation" is being used in a common way, so that explanations are context-dependent, and what constitutes a satisfactory explanation, like a satisfactory description, is relative to the purposes, interests, and background knowledge of speaker and audience.

expression (e.g., "animals belonging to *homo sapiens*" of "humans"). In the same way *certain* specifications of the sense of "(bodily) pain"—e.g., the quite unsatisfactory (E_1) (with a suitable choice of g) or the more satisfactory

(E_3) "(Bodily) pain" has the same sense as "sensation of a certain sort, viz., the distressing sort typically felt when physiologically hurt,"

can be consistently adjoined to a specification like (E_2) (with the same choice of g) of the total extension of "pains." But though we can consistently couple specifications of sense like (E_3) with (E_2) we don't have to. Therefore a contingent identity theory with a main hypothesis like (B) need not provide an analysis, topic-neutral or otherwise, of sensations or specifications of the senses of "bodily pain," "after-imaging," and so on.

(E_2) can be more explicitly formulated as

(E_4) The total extension of "SENGHS" is identical with the total extension of "PROGS."

The only material advantage of (E_4) over (B) is that it is clear that the sentence yields a contingent statement; and this could be made clear in (B) by replacing "identical" by "contingently identical" (call the result (B_2)). Unfortunately (E_4) leaves us with the same major problems as (B) or (B_2): for criteria for and conditions on identity are crucial. These criteria a sense/reference distinction does not automatically supply. Although it is certainly valuable to have a sense/designation distinction, or at least to have the tie-up between contingent identity and sense clarified, in order to reinforce the argument at various points, the subsequent discussion could be accomplished without appeal to such distinctions. Sense/reference distinctions can be by-passed.

The usual (Leibnitz') definition of *identity*,

- (1) $x = y$ FOR $(\forall f) \cdot f(x) \equiv f(y)$,
 where ' f '³ is a sentential function (or predicate),
 leads at once to
 (2) $[x = y] \supset \cdot (\forall f) [f(x) \equiv (y)]$,

the principle of indiscernibility of identicals; and to (3) $(\forall f) [f(x) \equiv f(y)] \supset [x = y]$,

the principle of identity of indiscernibles. The argument-range of ' x ' includes class- as well as individual-expressions. For almost all purposes, outside limited contexts like extensional logics where the class of predicates is restricted, (1) is much too strong. In particular, if adopted as a criterion of identity it would eliminate all scientific identifications and even many identifications within mathematics. For consider a tentative identity: $x_1 = y_1$, e.g., an identity of the sort made in scientific reductions like

(F_1) The temperature of an (ideal) gas is the mean kinetic energy of its molecules (strictly: $T = \frac{2}{3k}E$);

and consider the property specified by "is known by Arthur to be identical with x_1 " or abbreviated " $K_a(\dots = x_1)$," as in the statement "is known by Arthur to be identical with the temperature of a gas." If Arthur does not know much science we have as true statements $[K_a(x_1 = x_1)]$ and $[\sim K_a(y_1 = x_1)]$, whence by (1), $[x_1 \neq y_1]$ is true. There is an obvious way out of this impasse, as well as out of some standard difficulties raised by (3) over items with identical properties apart from spatio-temporal or temporal location in universes with limited accessibility, and by (2) once modal principles are introduced into predicate logics or set theories. First, limit the class of properties admitted and second, restrict conditions under which the definitions or implications hold. To achieve the first of these moves some preliminary terminology is valuable.

A property f is *intensional* or *non-extensional* iff either (i) ' f ' has a significance-range which includes the statement-expressions or infinitival expressions represented by ' p ' and ' q ' but f fails to satisfy the condition

(Ext) $(\forall p, q) \cdot p \equiv q \supset \cdot f(p) \equiv f(q)$

or (ii) ' f ' does not have such a significance-range but ' f ' contains a subpredicate ' g ' and g satisfies

³ Angle quotes represent the quotation function 'qu', which is so defined that for any given argument-value the value of the function is the quotation-expression of that argument. A statement-expression is an expression of the form "that..." where the blanks are filled by statemental sentences. Statemental sentences can also belong, derivatively, to the significance-range of ' f '. An infinitival-expression is an expression of the form "to..." where "to" is part of an infinitive, e.g., "to go home," "to be Pope." Some such expressions are needed if properties like "Tom wants..." "Tom liked..." are to be classed as intensional under the definition and grammatical ugliness such as "Tom wants that he is Pope" avoided. The definitions are phrased negatively in order that such properties as "... is tall," "... is blue," "lasts t seconds" may qualify as intensional, strict, etc. These properties qualify because their predicates fail to satisfy a significance condition. On approximately this point see S. Halldén, *The Logic of Nonsense* (Uppsala, 1949), pp. 47-48.

(i); otherwise f is *extensional*.^{3a} We abbreviate this definition: A property f is *intensional* if f fails on (Ext); otherwise f is *extensional* (abbreviated: *ext*). The definitions which follow are presented in a similar abbreviated form. A property f is *non-strict* if f fails on

(Str) $(\forall p, q) . \Box(p \equiv q) \supset . f(p) \equiv f(q)$;
and *strict* otherwise. A property f is *non-regular* if f fails on

(Reg) $(\forall p, q) . (\forall x) Kx(p \equiv q) \supset . f(p) \equiv f(q)$,
where ' Kx ' symbolizes " x knows (that)"; and *regular* (*reg*) otherwise. Thus " a knows . . .," " a believes . . .," and extensional properties like "is taller than a " are regular, but " a hopes" is non-regular. "It is necessary . . .," "It is possible . . .," and all extensional properties are strict, but "Tom believes . . .," "It is possible Tom wants . . ." are non-strict. A property f is *modal* iff f is strict but non-extensional; *narrowly propositional* iff f is regular but non-strict. Intensional properties include not only ". . . is known by a ," " a believes . . .," "is about a ," but—most important—" . . . is about something private," " a noticed . . .," " . . . belong to the surroundings of" and " a 's statement . . . is incorrigible." The non-extensionality of the last property can be shown by applying the definition to the equivalence:

[I have a pain] $\equiv$ [The man born in x_3 who has done f_3 has a pain].

Consider now the following identity definitions:

(4) $x = y$ FOR $(\forall \text{ ext } f) . f(x) \equiv f(y)$,

i.e., in classical logics:

$(\forall f) . \text{ext}(f) \supset . f(x) \equiv f(y)$.

^{3a} Clause (ii) of the definition of *intensional* is designed so as to include as intensional properties like "Arthur knows that Bill is the same individual as . . .," " a is frightened of . . ." For the relevant predicates contain non-extensional subpredicates; in the examples "Arthur knows" and " a is frightened" respectively. Strictly the definitions of *intensional*, *non-strict*, etc., are defective in the form given. To repair the defect 3-valued significance logic has to be used explicitly. The definition of *intensional* is repaired by replacing (Ext) by

(i) $(\forall p, q) . \Sigma^< f(p) > \& \Sigma^< f(q) > \supset . p \equiv q \supset . f(p) \equiv f(q)$.

To be exact similar amendments should be made to (Str) and (Reg). In these modified conditions ' Σ ' symbolizes "is significant" (Halldén's '+'), and ' $\equiv$ ' represents a 3-valued equivalence which is true if its components have the same value and false otherwise. Alternatively (i) can be replaced by

(ii) $(\forall p, q) . p \equiv q \supset . \Pi(f(p) \equiv f(q))$,

where $[\Pi p]$ is false when $[p]$ is false and true otherwise.

⁴ For, using Feys's system T (a system displayed in A. N. Prior, *Formal Logic*, 2nd edition [Oxford, 1962], p. 312) and classical predicate logic, it follows:

1. $[x = x]$
2. $\Box[x = x]$
3. $[x \equiv y] \supset [f(x) \equiv f(y)]$
4. $[f(x)] \supset [[x \equiv y] \supset f(y)]$
5. $[\Box[x = x]] \supset [[x \equiv y] \supset \Box[x = y]]$
6. $[x \equiv y] \supset \Box[x = y]$
7. $[\sim \Box[x = y]] \supset [x \neq y]$
8. $\Delta[p] \equiv \Diamond_{\text{pr}}[p] \& \Diamond[\sim p]$
9. $[\Delta[x = y]] \supset [\sim \Box[x = y]]$
10. $[\Delta[x = y]] \supset [x \neq y]$.

from (4) or any other plausible definition of ' $=$ '.

from 1, by the rule of T.

provided f is strict; from (5).

from 3, by propositional logic.

substituting ' $\Box[x = \dots]$ ' which is a strict predicate for ' f ' in 4.

from 2, 5.

from 6.

(definition of " p " is contingent).

from 8.

from 7, 9.

(5) $x \equiv y$ FOR $(\forall \text{ strict } f) . f(x) \equiv f(y)$.

Other identity criteria may be obtained by considering different classes of properties. All these identity criteria may be connected by the scheme for R-identity.

(6) $x =_R y$ FOR $(\forall f \in R) . f(x) \equiv f(y)$,

where R is a class of properties which includes at least all extensional properties. For example *extensional identity* defined by (4) is obtained by taking R to coincide with the class of extensional properties; *strict identity* defined by (5) by taking R to be the class of strict properties; *Leibniz' identity* defined by (1) by taking R to be the class of all properties.

Even strict identity is too strong if (B) or *any contingent* identifications are to survive. Consequently any identity criterion such as Leibniz' identity which entails strict identity is also too strong for contingent identities like (B). For if it is only a contingent matter that x is identical with y then x is not strictly identical with y .⁴ Since many sentences containing "identical with," in particular (D), (F₁), and (B), only yield synthetic statements strict identity is too strict. We must relax our requirements below those of Leibniz', and step down at least to a class R of properties from which most intensional properties are excluded. The example of Arthur shows that " $Ka \dots$ " must be excluded from R . Similar examples can be constructed to show not merely that all narrowly propositional properties but that very many other intensional (mental) attributes have to be excluded from R . Further examples can be designed

to show that many impure properties (those whose specifying predicates include essentially quotation) and [as in step 5, footnote 4] many modal properties which can be significantly attributed to individuals fail to qualify for R . Because only extensional properties clearly qualify for R , because of the indeterminacy of the class of further properties which might qualify for R , and because (4) is sufficient to retain important substitution (*salva veritate*) principles and sufficient to guarantee identity of reference, we propose to step down to (4). (4) certainly seems to be very like what is required for a satisfactory contingent identification; any criterion much stronger would bring down most scientific identifications.

If Leibniz' or even strict identity were taken as a criterion of contingent identity (symbolized ' $\cong$ ') it could be shown

$$x \cong y \equiv \Box(x \cong y)$$

and using Lewis' system S5 we obtain

$$\sim(x \cong y) \equiv \Box \sim(x \cong y).$$

Thus a contingent identity would be logically impossible. All true identity and difference statements would be analytic. Such a result is but a special case of the principle that all true statements are analytic, from which Leibniz claimed to deduce (1). Identity criteria (1) and (5) can be retained only if either all identities are analytic or else all predicates demarcate extensional properties. Both alternatives are false. Therefore (1) and (5) cannot be retained as contingent identity criteria. An identity criterion syntactically identical with the Leibniz criterion can, however, be obtained by pruning down the class of properties. Instead of redefining the word "property" let us call the properties which are so admitted *traits*. Many predicates which demarcate properties will not demarcate traits. So an important question is: Which properties are traits? If traits are identified with extensional properties, then (1) holds provided the *interpretation* is changed so that ' f ' now takes as values not properties but traits. Such a procedure, to keep up old appearances, provides only a thinly-disguised terminological variant of the course we have followed.

There are other ways of *apparently* avoiding replacement of (1) by (4), *extensionalizing moves*, which propose replacement of all intensional properties either (i) by (distantly related) extensional properties or (ii) by extensional and impure properties. There are also connected extensionalizing

techniques, used to retain (2), which come down to allowing substitutions only in extensional sentence contexts. But serious objections to both (i) and (ii) seem unlikely to be surmounted. Moreover (i) immediately concedes (4) for *all* identities and thus it strikes difficulties which adoption of (4) just for extensional identity avoids. For example, the important distinction between contingent and strict identity becomes difficult to draw, and escape routes from the Black dilemma [see below] are blocked. In case (ii) Leibniz-identity is qualified in the usual manner so as to exclude sentential functions which contain quotation essentially. If (ii) were viable this would include all intensional properties not already replaced by extensional analogues. Therefore (4) is conceded for all identities. Consequently (ii) encounters other difficulties, one of them being the problem of how to define non-extensional identities.

Direct arguments for criterion (4) run like this: If x and y are extensionally identical then they share all extensional properties. What needs to be shown, then, is that non-extensional properties are irrelevant to extensional identities. If x and y are extensionally identical it is only the extensions of ' x ' and ' y ' which are relevant in the identification. But non-extensional properties, such as modal properties, concern the sense of ' x ' and ' y '. Therefore extensional identities are independent (in the sense explained earlier) of non-extensional properties; and (4) is established as an appropriate criterion for extensional identities. Lastly (4) can be made plausible by examples: consider sentences (G1)–(G5) below.

Given the adoption of (4), it follows that 'iff' in (C)-sentences only symbolizes a bi-conditional of approximately the same strength as a truth-functional bi-conditional and of much less strength than a mutual entailment or strict equivalence; and, given also classical two-valued logics, 'iff' does represent truth-functional equivalence. An instantiation of (C₂), namely

$$(C) \text{ [SENGH } (x_0)] \equiv [\text{PROG } (x_0)],$$

follows at once from the set-theoretic formulation of (B).

Adopting (4), i.e., opting for *extensional identity* at once eliminates several standard objections to (B); e.g., that various people can know about or understand details of one process but not the other, that the hypothesis is not good enough because, since the identity is only a contingent

matter, it could be false, that SENGHS are private but PROGS are not, that SENGHS are necessarily mental but PROGS are not, that SENGHS are sometimes noticed when PROGS are not, and that SENGHS require surrounding intensional practices and assumptions but PROGS do not. Adopting (4) also eliminates the Black dilemma⁵ constructed on the assumption that SENGHS are identical with PROGS: either there is no way of distinguishing SENGHS and PROGS because they are identical or there remain some irreducible mental properties. First, one horn of the dilemma breaks down once (4) replaces (1). SENGHS and PROGS can be distinguished even though identical using intensional properties. Since not all intensional properties are mental properties, mental properties are not indispensable in separating SENGHS from PROGS. Second, there is nothing in this separation to stop *extensional* identifications, e.g., of beliefs with physiological states. (B) is neutral with respect to contingent identifications, and even with respect to certain analyses of narrowly propositional properties in physicalistic, behavioristic, or other styles.

There are further difficulties which the adoption of (4) does not forestall; difficulties which face any *heterogeneous identification* or *reduction*. There are essential differences between

(I) *homogeneous identifications*; e.g., (D) or:

- (G₁) Scott is (identical with) the author of Waverley.
- (G₂) The Morning Star is (identical with) the Evening Star.
- (G₃) The successful general is the same person as the small boy who stole the apples.
- (G₄) $7 + 5 = 12$.

To be sure there are *important* differences between (D)s and (G)s on the one hand, and individual members of each of these collections on the other—and also between these and

(II) *heterogeneous identifications*; e.g., (F₁) or:

- (F₂) Lightning is an electrical discharge from ionized clouds of water-vapor to the earth.
- (F₃) Material objects are certain structured collections of atomic particles.
- (F₃¹) Rubber (caoutchouc) is atoms of carbon and hydrogen structured in polymers of isoprene.

⁵ This dilemma comes from an objection raised by Max Black to Smart's identity theory: see J. J. C. Smart, "Sensations and Brain Processes," *The Philosophical Review*, vol. 68 (1959). The objection is set out more explicitly in J. Shaffer, "Mental Events and the Brain," *The Journal of Philosophy*, vol. 60 (1963), pp. 160–166. For standard objections to identity hypotheses which adoption of extensional identity, or of contingent identity, knocks out, see N. Malcolm, "Scientific Materialism and the Identity Theory," *The Journal of Philosophy*, *ibid.*, pp. 662–663; Smart, *op. cit.*; Shaffer, *op. cit.*; J. Shaffer, "Recent Work on the Mind-Body Problem," *American Philosophical Quarterly*, vol. 2 (1965), pp. 81–105, and J. J. C. Smart, *Philosophy and Scientific Realism* (London, Routledge & Kegan Paul, 1963).

- (F₄) Visible light is electromagnetic radiation of wavelengths between ca. 4000 and 7200 Å.
- (F₅) Pleasure is a certain sort of tension reduction.
- (F₆) Real numbers are classes of rational numbers.
- (F₆¹) $\sqrt{3} = \{x: \text{rat } (x) \ \& \ (x^2 < 3)\}$
- (F₇) Cardinal numbers are classes of classes.
- (F₇¹) 3 is the class of all classes which correspond 1–1 with the triple (x, y, z) .

The defining difference between (I) and (II) is sometimes explained in this way: the levels or types of the referring expressions on each side of any selected identity sentence in (I) are the same, while in (II) they are different. But it can be explained more satisfactorily thus: in (I) the same predicates and sorts of predicates can be significantly predicated of the referring expressions on each side of any selected identity-sentence, but in (II) *some but not all* predicates can be significantly predicated of both sides of an identity sentence.

As an illustration consider

- (F₈) the only object on the top of Everest = the flagpole erected by Hillary.

Since the requisite object might have been Hillary, and since significance is context-independent, it is significant to predicate of the left-hand side (LHS) of (F₈) but *not* of the right-hand side (RHS) of (F₈) such predicates as "was born in New Zealand," "likes ice cream," "is thinking about Vienna." Any number of examples like (F₈) can be concocted by taking descriptions of different types; as examples consider "The only thing ever to have scared Smith is the Benfield apparition," "The white dot on the horizon is Austin's house" and (F₉): the intersection of the classes of points described by the linear equations $(x = y)$ and $(x + y = 2) =$ the point with coordinates (1, 1).

If points are individuals, it is significant to predicate of the LHS but not of the RHS of (F₉) "has one member" and "contains one point." Examples designed to show that (F₁)–(F₇) are properly classed as heterogeneous are, like most significance claims, controversial. We suggest: "Is the mean of the same property of its microcomponents" is significantly predicable of the RHS but not of the LHS of (F₁), and "is felt by most humans" and

"occasionally fogs up the windows" of the LHS but not of the RHS of (F_1) . "Struck Tom," "is roughly chain shaped," "is regularly accompanied by thunder," and perhaps "is caused by an electrical discharge" and "is yellowish-blue," are significantly predicable of the LHS but not of the RHS of (F_2) , and "is sometimes elliptically polarized" of the RHS but not of the LHS of (F_2) . "Get broken," "are sat on," "are sawn into little bits," "are glued together," "have solid parts," "are of uniform material throughout," "contain no space" of the RHS but not of the LHS of (F_3) . "Has many members" of the RHS but not LHS of (F_7^1) and (F_6^1) . And so on.

(B)-hypotheses are heterogeneous identifications. "Contain a negative feed-back loop," "are detectable indirectly upon cutting open a cat's skull," "travel along nerve fibers," and "occur in brains," are significantly predicable of the LHS but not of the RHS of (B) and (B_1) ; and perhaps "hurt" and "are exhibited in external behavior" not of the LHS of (B) and (B_1) .

Not only (B)-hypotheses but identities like (F_8) and (F_9) and well-confirmed and accepted theoretical identifications like (F_1) and (F_2) , indeed all heterogeneous identifications, flounder when extensional identity is adopted as identity criterion. For how can a and b be identical if there is an extensional property f_0 such that " $f_0(a)$ " is significant but " $f_0(b)$ " is not? a and b would not have all their extensional properties in common, and so would fail to be identical even under the extensional criterion. A way to avoid these disasters, which has considerable appeal for other reasons as well, is to replace (4) by (7):

$$(7) \ x \cong y =_{\text{Df}} (\forall \text{ ext } f) . \Sigma^<f(x)> \ \& \ \Sigma^<f(y)> \supset . f(x) \equiv f(y);$$

to adopt what we call *contingent identity* or *co-incidence* as defined by (7) as at least a *necessary* condition for the truth of heterogeneous identification statements. Call this condition "*requirement (a)*" on contingent heterogeneous identification. More generally, (6) is replaced by

$$(8) \ x \cong_R y =_{\text{Df}} (\forall f \in R) \Sigma^<f(x)> \ \& \ \Sigma^<f(y)> \supset . f(x) \equiv f(y),$$

where ' Σ ' symbolizes the predicate "is significant," a predicate whose significance-domain includes all quotation-expressions. In the theory of significance presupposed there are always predicates, e.g., "is an item," "is an occurrence," "is had by a ," "lasts

time t " in the case of (7), such that the antecedents of the definiens of (7) and (8) are satisfied.

When an identification is homogeneous (7) reduces to (4) if non-significant sentences are mapped uniformly into false statements, or else uniformly into true statements. A similar move would not rescue (4) in the case of heterogeneous identifications. If non-significant predications were mapped into true attributions then all identity statements like (F_8) and (F_9) (where negative significance-ranges of LH and RH expressions differ) would come out false; while if non-significant predications were mapped into false attributions all identity statements like (F_1) and (F_2) (where positive significance-ranges of LH and RH sides differ) would be swept away.

A favored manoeuvre, made to avoid replacement of (4) by (7) and sometimes to avoid replacement of (1) by (4), is to redefine refractory predicates so that they can be transferred from one side of an identity sentence to the other preserving truth-value. Suppose

$$(a) \ [x = y] \text{ is true, but } \Sigma^<f(x)> \text{ and } \sim \Sigma^<f(y)>.$$

Then ' f ' is so redefined and/or ' y ' is enriched by such conventions that at least

$$(b) \ [f(x) \equiv f(y)] \text{ is true.}$$

But, first, the manoeuvre presupposes (7). Allowing redefinition or the consistent addition of conventions here is tantamount to recognizing significance requirements. For if (7) is not correct the supposed case (a) is impossible: how, otherwise, could $[x = y]$ be true if predicates like ' f ' were not discounted? The same points tell against any attempt to save (1) by redefining modal predicates so that they transfer across contingent identity sentences. Allowing such redefinition amounts to discounting or dismissing the initial properties. Second, without (7) how is it to be decided which predicates can be non-viciously redefined? How can predicates whose redefinition is permissible be distinguished from predicates whose properties *falsify* the identity statement? Suppose that redefinitions of predicates (and referring expressions) are proposed without the introduction of (7). Then the redefinition ploy is vicious and very likely circular. To make appropriate redefinitions the identity should be known to be true. But if the identity is true redefinition would be otiose as far as guaranteeing the identity goes. If redefinition is needed to bolster up the identity then it must be in some way defective: how then can it form the

basis of appropriate redefinitions? Unless the identity statement is true and its truth determined independently using an identity criterion like (7) it cannot be made a suitable basis for redefinition. If defective identities could be made the basis of redefinition there would be serious danger both of rendering apparently correct defective identities and of guaranteeing the correctness of an identity only at the cost of its trivialization.

To stress these important matters consider the full materialism:

(M_1) : minds = brains

since it is much more vulnerable than (B)-hypotheses. For there are predicates like "are spatially locatable" and subpredicates of "are material things" like "are colored," "are not heavy" which are significantly predicable of the RHS of (M_1) , but what of the LHS? Redefinition of these predicates so that a "mind" has the location, color, and weight of its correlated brain not only begs the question since the truth of (M_1) is assumed in the definitional extensions of the predicates but also (what is in general inadmissible) changes the sense of "mind." If the sense of "mind" is sufficiently amended, in the limit to "brain," (M_1) will certainly be consistent: but trivially so. Worse, all the difficulties of extensionalizing moves are then encountered. In particular, the identity statement (M_1) is *not synthetic* if the sense of "mind" is sufficiently amended. Also, then, the Black dilemma and related puzzles are unavoidable.

A vital question, if redefinition and extension of significance ranges does not always trivialize an hypothesis, is: When is redefinition admissible? At least when (a) occurs and the significance range of ' f ' can (significantly) be extended as in (b) on the basis of the true identity $[x \cong y]$, where the truth of $[x \cong y]$ is assessed by using (7); but not in general. If correct, it follows that redefinition of predicates which are not impure simply to render an hypothesis consistent is inadmissible; for the truth, and hence consistency, of the hypothesis is assumed for redefinition. It also follows, since $[f(y) \supset \Sigma'f(y)']$, that the significance-range of ' f ' is extended under redefinition and that the sense of ' f ' is revised. Such (conceptual) revisions *do* occur on the basis of identity statements. Consider the extensions of significance-ranges of color predicates to include expressions referring to certain electromagnetic phenomena. Using our theory based on (7) these revisions can be comprehended.

Replacement of (4) by (7) at once eliminates those criticisms of (B) based on the existence of properties of physiological processes, the predicates of which properties are not significantly predicable of sensation-expressions, and mitigates those criticisms based on the presence of properties of sensations whose predicates are not significantly predicable of physiological referring expressions. To destroy this last criticism altogether requirement (β) on heterogeneous identifications [see below] has to be met. Roughly, certain emergent properties of sensations have to be explained in terms of the reducing physiological theory much as, in the case of (F_3) , macro-properties of material objects such as temperature and solidity (in the relevant sense of "solid") are explained under the reducing theory in terms of microstructures. In other words, replacement of (4) by (7) eliminates or softens several criticisms directed at (B)-hypotheses because they are heterogeneous identifications. The criticisms are thwarted because to establish contingent identity it is sufficient to show, for every extensional property f whose predicate is significantly predicable of SENGHS and of PROGS:

$[f(\text{SENGHS})] \equiv [f(\text{PROGS})]$.

Now restrict consideration to synthetic identifications; so mathematical examples can be discarded. Though there are grounds for distinguishing structural and state identities from process and event identities we shall not develop such distinctions. In fact, but not necessarily, all remaining identifications in (I) are observational identifications made from observational bases largely on grounds of spatio-temporal continuity and coincidence, non-separability, non-independent manipulability, and preservation of observational features. (Puzzles of personal identity and about sameness of boats and cars under various changes eject most criteria for sameness of actual referents in homogeneous identifications.) These grounds are also important in theoretical identifications falling under (II) but they are *not* nearly *sufficient*. For one thing, requisite observations may be physically impossible or seriously limited. The scientific identifications made under (II) exhibit a certain asymmetry. First, one referring expression in each, the one written typically on the right in English, is a theoretical expression explained (partly) within the context of a certain theory, while the other expression is usually an observational expression. Second, the identification is explanatorily directed; the item referred to by the

right expression is designedly introduced or used to explain the item referred to by the left expression and features of the item. This helps elucidate the point of the dictum that the "is" in heterogeneous identifications is the "is" of explanation. The dictum is misleading because there is not a special sense of "is" used just in heterogeneous identifications and because explanatory power is not a sufficient basis on which to make heterogeneous identifications. Moreover, if "Lightning phenomena can be explained by the theory of electrical discharges" were all that "Lightning is an electrical discharge" meant, we could still ask about the relation between lightning and electrical discharges. Because of the asymmetry we can distinguish LH and RH expressions as *reducenda* and *reducentes* respectively. The identification sentences are *reduction sentences* and the identity expressed in each is the *reduction relation*. A *heterogeneous reduction* is a reduction where there are predicates significantly predicable of the *reducendum* which are not significantly predicable of the *reducens*; or, where a whole theory or a region of discourse and truths and well-founded or accepted assumptions and discriminations characteristically expressible in its vocabulary (a conceptual framework?) is being reduced, a reduction where the reduced theory includes in its formulation or vocabulary distinctive (descriptive) predicates which are not included in (initial) formulations of the reducing science⁶. The class of heterogeneous reductions overlaps that of heterogeneous identifications. (B)-sentences formulate heterogeneous reductions and theoretical identifications. That many expressions referring to sensations, like "Tom's toothache," "My giddiness," are observation expressions, and that many of the requisite neurophysiological terms needed in the identifications will be theoretical, though both controversial, seem to us defensible claims.

That (a) is not a sufficient condition for a theoretical identification is in part a consequence of the requirement that the reduced phenomena (or items) and further properties of the reduced phenomena be explained by the reducing phenomena (and the theories in which it is treated). The *reducens'* referent cannot provide the requisite explanation on its own. The identification is not made independently of the theory to which the *reducens* belongs. It is the availability of such a *theory*, containing law statements in terms of which

the reduced phenomena can be explained and their further related properties explained or predicted (perhaps using also associated theories), that provides the *further* grounds or case for the identification. If it were not for the theoretical connections the (open) universal statements yielded by sentences of (II) could not be adequately tested or confirmed. The further case for adding an explanatory requirement is this: With theoretical identifications several of the usual tests for identity, those used in observational identifications, are ineffectual or inapplicable; further the class of putative common properties is considerably reduced. To compensate for this diminution of requirements and to guarantee the dependence and concomitant variation of the identified items some further requirement is needed. A suitable requirement, which preserves the asymmetry of heterogeneous identifications, is the explanatory requirement. Lastly, in order to decide between rival verificationally equivalent hypotheses [see Sect. IV] further conditions on identity are needed.

An explanatory requirement (β) constitutes a necessary condition for the truth of a theoretical identification. Let ' x ' range over *reducenda*, and ' x_0 ', ' x_1 ' . . . be *reducenda*, e.g., expression referring to sensations. Let ' y ' range over *reducentes*, and ' y_0 ', ' y_1 ' . . . be *reducentes*, i.e., expressions referring to reducing items. Let the identity be formulated, generally, " $w \cong v$ ", where ' w ' designates the class of reduced items, i.e., the class of x 's and ' v ' the class of y 's. Then we require:

(β): There are scientific theories T_1 and T_2 such that y -expressions belong to the vocabulary of T_1 , and such that T_1 and T_2 —possibly together with experimental data about y 's and other items treated in T_1 —jointly explain both x 's and extensional (T_1 & T_2)-linkable properties of x 's whose predicates are significantly predicable of y -expressions, and also explain some extensional (T_1 & T_2)-linkable properties of x 's whose predicates are significantly predicable of x -expressions but not of y -expressions. A property f is *T-linkable*, i.e., theory-linkable under theory T , if f is not a relational property relating those items of which it is a property to items whose attributed behavior is not within the scope of T . For example, frightening many females is a relational property of lightning but not a property we expect electromagnetic theory to explain.

To illustrate: in order that hypotheses (B) be

⁶ See E. Nagel, *The Structure of Science* (London, Routledge & Kegan Paul, 1961), p. 342. The "homogeneous-heterogeneous" terminology has as sources the cited text and the work of B. Russell and of L. Goddard on significance-theory.

established there must be formulated a theory P_1 of physiological processes which together with more comprehensive theories P_2 , e.g., chemical and electromagnetic theories and possibly quantum theory, explains all those extensional (P_1 & P_2)-linkable properties of sensations whose predicates are significantly predicable of expressions referring to physiological processes, e.g., those relating to duration, pattern, intensity, location of sensations, and most (ultimately all) of those genuine extensional (P_1 & P_2)-linkable properties of sensations whose predicates are not significantly predicable of physiological processes.

(β) is a powerful requirement. The truth of statement (β) entails that all the standard *requirements on scientific explanations are satisfied*,⁷ which in turn entails that the usual requirements on a scientific reduction, insofar as they are applicable to sensations (since we do not have a scientific theory of sensations) are satisfied.

Finally (α) and (β) are jointly sufficient conditions for heterogeneous theoretical identifications or reductions like (B). If only a very limited class of sensations, say tickles, were identified with physiological processes then it would be a fair criticism that it is also necessary that further items, e.g., further features of *other* sensations be explained as well as those of tickles. When, however, occurrences as varied as the various sorts of sensations are explained and connected, the further necessary condition envisaged is automatically met.

It seems to us that (B)-statements, or at least statements yielded by (B)-sentences when the variable-places are filled in, can *consistently* satisfy requirements (α) and (β). We can, in fact, already go some little distance toward outlining general features of a requisite theory P_1 . Whether, however, such a theory when developed will contain or permit the expected reduction of sensations, seems at present open. If our surmises are correct (B)-hypotheses are *viable but only hypotheses*.

III

A major objection to combat is: (B)-statements cannot consistently satisfy a contingent identity criterion. Most discussion of (B)-type hypotheses

has, in effect, centered either on whether this major objection is correct or on whether (B)-hypotheses are empirical and on means of deciding between (B)- and allegedly rival hypotheses. The second of these issues we reserve for (IV) below; to the first we now turn.

It is argued: there are predicates, satisfying the antecedent of (7), such that the properties they specify are properties of SENGHS but not of PROGS; most notably the pairs "public," "private" and "corrigible," "incorrigible." We have pointed out that the properties specified by "is incorrigible to a " and "is about something private" are non-extensional, and thereby excluded. Suppose, in reply, exception is taken to the extensionality restriction itself, that it is objected to as being much too restrictive, that it is contended that the class of properties admitted under (7) should be widened to include other properties like incorrigibility. Arguments independent of (7) can be advanced for excluding such properties; e.g., comparisons of giddiness zones and pain patches with arid zones and brown patches can be used in arguing that being private and incorrigible are different *sorts* of properties of sensations from lasting t seconds or regularly exhibiting an α -rhythm and are not relevant to scientific identifications. If, however, we were to play along with the opposition the way would be open for the launching of an argument like the following:⁸

1. [$'p'$ is a first person sensation report (statement)] implies [$'p'$ is an introspective report]
2. [$'p'$ is an introspective report] implies [$'p'$ is wholly about something private]
- ∴ 3. [$'p'$ is a first person sensation report] implies [$'p'$ is wholly about something private]
4. [$'p'$ is wholly about something private] is incompatible (inconsistent) with [$'p'$ is about something public]
5. [$'p'$ is a physiological (physical) report (statement)] implies [$'p'$ is about something public]
- ∴ 6. [$'p'$ is a first person sensation report] is incompatible (inconsistent) with [$'p'$ is a physiological report];

from 3, 4, 5 using propositional calculus (or modal logic).

⁷ On these requirements, so far as they are worked out at present, see C. G. Hempel and P. Oppenheim "Studies in the Logic of Explanation," *Philosophy of Science*, vol. 15 (1948), p. 135; C. G. Hempel, "Deductive Nomological vs. Statistical Explanation," *Minnesota Studies in the Philosophy of Science* (ed. H. Feigl and G. Maxwell), vol. 3 (Minneapolis, University of Minnesota Press, 1962); and E. Nagel, *ibid.*, pp. 30-46.

⁸ This argument and the set of logically necessary conditions for *privacy* are extracted from K. Baier, "Smart on Sensations," *Australasian Journal of Philosophy*, vol. 40 (1962), pp. 57-70. The argument is a reconstruction of Baier's main argument against Smart.

The argument is valid but several premisses are either suspect or defective. A combination of the following points is sufficient to destroy the argument. First, it is misleading to call first person sensation reports “physiological reports,” because an identity of sense is suggested. The employment of the indeterminate word “about” does not help in rectifying this fault. In the requisite sense of “about,” in which two different expressions are *about* the same thing and where contingent but not strict identity obtains, ‘ x_0 ’ can be both about something private (a sensation) and about something public (a physiological process). Second, premisses 2, 3, and 4 are false; they depend on a neat but questionable public/wholly private dichotomy. The usual senses of “public” and “private” (v. OED) cannot be used to support the dichotomy or in defence of 3, 4, and 5. Thus to back the argument technical senses of “about something private” and “is private” must be introduced. Consider the following fairly typical dualistic set of logically necessary conditions (with a brief justification added parenthetically) for the truth of “ $f(x_0)$ is about something private” or “ x_0 is private”:

- (a) x_0 is owned, i.e., is had by someone.
- (b) x_0 is exclusive or unsharable (because two cannot share *one* pain).
- (c) x_0 is imperceptible by the senses (for “ a sensed [e.g., saw, felt, tasted] a ’s pain,” “ a sensed b ’s pain” are both non-significant).
- (d) x_0 is asymmetrical with respect to persons (for “I could sense [e.g., see, hear] I had a pain” is non-significant but “I could sense he had a pain” is significant; and x_0 is something about the having of which the haver could not, but other people could, significantly consider or weigh evidence; e.g., “I am weighing the evidence for my being really in pain” is non-significant).
- (e) x_0 is such that the person who has it has final epistemological authority about it (for “I have a pain unless I am mistaken” is non-significant. The authority is based on the haver’s being the person best placed to discover his mistake or confirm his belief).

These conditions are not only an inadequate base from which to spring to premisses like 3, 4, 5, but indeed unsatisfactory on other counts. First, observe that there are types or sorts of sensations as well as token-sensations. A type/token distinction can be readily transferred or applied to processes and occurrences. In fact many physiological processes and a good many sensations have been typed or classified. Token-sensations are (almost)

always had, experienced, undergone by an *individual* person, they can be assigned a given time bound, and they do not recur even though they may be a recurring type of sensation and have recurrence patterns within the time bound. Often, however, we refer to type-sensations; then a person may have the same sensation, e.g., a certain rheumatic pain, at widely repeated intervals, and several persons may have the same sensation, e.g., the same after-imaging, the same sort of, or the same, thirst or giddiness. Now token-brain-processes satisfy both (a) and (b); for they belong to and are in fact had (even if “felt” is not significantly applicable) by one person and they are exclusive, except possibly in cases like Siamese twins which create special difficulties about token-sensations too. One effect of (b) is to restrict consideration to token-occurrences: type sensations are not private. Against the comeback that it is not logically necessary that token-brain processes are owned and exclusive two moves are feasible: either to argue that this doesn’t matter because (B)-statements are only synthetic; or to query the comeback and point out that any argument to this effect can be matched by an analogous argument in which “token-sensation” replaces “token-brain-process.” Next, (c) seems to be mistaken unless qualified; pain-receptors play an important part in the “perception” of pain, e.g., when hit with a cricket ball. Moreover “ a felt b ’s pain” is significant, unless “token-pain” is so defined that a token-pain is unsharable (as suggested in (b)). Such a definition is unpersuasive: if a ’s g -nerve were connected to b ’s g -nerve, through which pulses correlated with b ’s brain were being transmitted by a machine which did not interfere with pulses in b ’s nerve and which produced a pulse pattern (taking account of feedback effects if any) in a ’s nerve exactly similar to that in b ’s then we should say, probably along with most English speakers we guess, that a felt b ’s pain, i.e., one and the same pain. Incidentally, it can be admitted that one’s own pain-experiences are perceptible in a much more direct way than any physiological processes, and that they are perceptible in a more straightforward way than another’s pain-experiences without damage to identity theses. And, relying on (7), it can be conceded that while, e.g. “ a feels his own pains” is significant “ a feels his own brain processes” is not. The cogency of condition (d) depends on how “asymmetrical” is elaborated; for being sweaty under the arms, or having a mouth ulcer, is asymmetrical but not

wholly private. Elaborations suggested by the supporting examples should be jettisoned; for "I felt I had a pain," "I thought I had a pain," "I am debating whether I really have a pain" (I have a toothache) when the putative pain is of low intensity, and the seldom necessary "I can jolly well see I'm in pain, not just feel it" when remonstrating with someone, are significant, and the statements yielded may be true. Also a person may well have to weigh evidence during a sensation to determine what sort of sensation it is, or after an occurrence during which the weighing of evidence is out of question. Finally, it is not analytic that a person is the final epistemological authority on his sensations: (e) is too strong if genuine pains are to rate as *private*. A person is only the final authority on when he thinks or sincerely supposes he has a sensation; and having a sensation or having a genuine sensation *differs* from thinking that one has a sensation, and this from sincerely reporting that one seems to have or thinks one has a sensation. He is the final authority on when he thinks he has one because "if *a* sincerely supposes he has a pain then he has a supposed, or imagined, pain" is analytic. Admittedly a person is the authority on his own pains, because he feels them. Admittedly "*a*, in general, knows better than others when he is in pain" is true: only the statement is not analytic, and the "in general" cannot be elided. When neurophysiological theory is much better developed than at present we expect that public and physiological tests for sensations will be stressed much more; and that in some cases the results of these tests will override a person's reports. Compare the evolution of different tests for the comparison of temperatures and weights.⁹

Since pain-occurrences present as tough a front as any sensations in respect to privacy and incorrigibility let us, in our further defense, concentrate on them. To the questions "Does the sentence 'If I feel a pain then I have a pain' sometimes yield a synthetic statement?" and "Is there room for the expression 'I seem to feel a pain' as distinct from 'I do feel a pain'?" we propose to answer yes. We submit that the English word "pain" stretches, or ranges, *between two contrasts*, first a (seeming) pain/no (seeming) pain at all contrast, and second a real pain/imaginary or imagined pain contrast. The contrasts dovetail for (seeming) pains are

either real or imagined pains. Though the "real pain," "imaginary pain" terminology occurs in several English idiolects,¹⁰ the basis of the distinction is not so clear. Often, however, the distinction seems to be made on physiological grounds; consider real toothache, genuine pain. We propose to adopt physiological grounds as necessary conditions for real or genuine sensations; a person has a real sensation, an *r-sensation*, only if he is in the appropriate physiological state. Appropriate state: for if an identity hypothesis proves correct imaginary pains are coupled or identifiable with *some* neurophysiological basis. What people seriously say will have a good deal to do, initially at least, with which physiological states are appropriate. We call the pains of the larger class, seeming or sensed pains, which are contingently connected with sincere reports, *s-pains*. It follows from the distinctions that feeling an *s-pain*, or seeming to feel a pain, is not the only criterion for actually feeling or having a pain, or for having an *r-pain*. Consider especially low intensity *s-pains* where one may check for tissue damage, tooth decay, etc. Or suppose *a* reports that he has a pain in his foot. Then *a* has an *r-pain* if he has the appropriate pain in his foot, a matter which can be checked, ideally, by two coupled dolor-meters, one on the appropriate nerves in his foot or toward the extremity of the relevant nerve in the case of amputations, and the other connected to probes in the appropriate area in *a*'s pain center. He has only an *s-pain*, if an identity hypothesis proves correct, if a pain-detector scanning neural sensation-areas for pain and sensations like pains responds; and his report is sincere only if as a matter of fact the scanning detector responds. Given the distinction between *r-pains* and *s-pains* proposed answers to the initial questions can be vindicated; for "If I have an *s-pain* then I have an *r-pain*" is not analytic since the statement is sometimes false; and "I feel a real pain" is distinct from "I seem to feel a pain," i.e., from "I feel an *s-pain*." Contingent identity hypotheses are, however, *consistent* with other answers to the initial questions than those we proposed. For if such hypotheses are correct *situations will not in fact arise* where relevant physiological data and sincere sensation reports conflict, even though it is logically possible that such situations occur. If a repeatable situation were to arise where someone

⁹ See also the weight analogy in J. Wisdom, *Other Minds* (Oxford, Basil Blackwell, 1956), pp. 101–105.

¹⁰ An empirical claim which we have so far only checked over a small sample. Were it not for the apparent ubiquitousness of "real," "imaginary" terminology we should have chosen less controversial and ambiguous expressions; for "real" can mean, in these contexts, "acute."

felt a pain but the pain-meter didn't move though it seemed to be working properly, then we should have to abandon or modify the theory after sufficient rechecking. In this respect there is not more difficulty in testing the identity hypothesis, *once* we have the appropriate and calibrated instrumentation [within the limits of verificational equivalence; see IV], than with most scientific theories—provided the hypothesis is openly falsifiable. The hypothesis may, however, at a later stage, having passed initial and repeated testing, cease to be openly falsifiable; it may become normic, i.e., nearer-analytic, and be adopted as *one* criterion for, or strongly-weighted mark of, a sensation. Physiological grounds would then be adopted as one criterion, and might even override apparently sincere reports of the person reporting that he has, or does not have, a given sensation. We might, under some circumstances, refuse to say that a person apparently sincerely reporting an after-image of a certain shape, size, and color was reporting correctly (if his visual apparatus, etc., was normal) if his report conflicted with the theory's predictions in respect to shape, size, or color. But if occurrences of this sort were frequent and varied enough we should be led to modify the theory. Even if such cases were to occur it would *not* follow that the sense of "I have a pain," etc., *must* have changed or shifted away from usual senses. For cases even now occur when a person may sincerely report a pain but many English speakers, at least, are prepared to say that the person did not *really* feel a pain, that he only *imagined* or thought he had a pain, or that he had an imaginary pain.

We outline documented support for the real/imaginary pain distinction. On October 23, 1963, in Sydney a woman was awarded Workers' Compensation by the full Supreme Court on the strength of delusional pains she had which were produced or contributed to by an accident at work. Both the judges and the ABC news reporter used the expressions "imaginary pains" and "delusional pains"; and the ground on which the pains were so-called was that there was no physiological basis for such pains.¹¹ It could be argued, though none too convincingly, that the expressions simply meant that she had no obvious damage or diseases rather than that there were no appropriate physiological processes *or* that she wasn't in real pain. None-

theless the case certainly seems to show that there might be situations where it is of interest and importance to use physiological criteria for *genuine* pains.

Many arguments alleging that (B) cannot consistently satisfy a contingent identity criterion lean heavily on the supposed incorrigibility of first person pain reports. But such reports are not absolutely incorrigible (to set out points summarily) because (i) a person can be mistaken and corrected as to whether he has *r*-pains. If we consider from now on only *s*-pains, then indeed a person can't be corrected if he *succeeds* in sincerely making the statement "I have an *s*-pain"; but only because if he sincerely reports that he has a pain, then he has an *s*-pain. And his sincerity may be questioned. Moreover, (ii) a person's *s*-pain statements may be misdescriptions. It is a mistake to dismiss misdescription and misclassification along with slips of the tongue, etc., as simply verbal errors. For in misdescription these errors are reflected in wrong or false statements, statements which might have quite serious non-verbal consequences. And the misdescription may result from lack of knowledge or mistaken beliefs about the pattern or features of his sensation; in which case it is not merely his words but his *beliefs* which are incorrect. Under the same heading we may include mistaken comparisons, which may occur, because of slips or, more important, because of mistaken assumptions or beliefs about language or about non-verbal matters. (iii) "Incorrigible" associates with "correct" and "incorrect." These words have multiple roles: "incorrect," for example, is highly determinable and covers a wide variety of criticisms; consider the sorts of items which may be called "incorrect" and the various different reasons for the incorrectness. Part of the guile of the incorrigibilist is not to disclose which reasons or criticisms he is prepared to take into account as relevant. "Incorrigible" can associate with "impossible"; incorrigibility should be concerned with the *logical* impossibility of correction of the speaker. In less stringent senses of "incorrigible" first-person pain reports are virtually incorrigible, e.g., there is little *real* possibility that the speaker will be corrected. (iv) The endeavor to rule out unfavorable examples by progressively tightening "incorrigible" so as to eliminate, along with lying and

¹¹ Two key passages read as follows: (i) (Mr. Justice Moffit in his judgment) "It was accepted as a fact that the applicant has a delusional state causing her to experience the symptoms of pain referred to," and (ii) (Sydney *Morning Herald's* Law Reports, October 24, 1963): "... this psychosis had produced after the accident a delusional state giving the applicant an impression that she was suffering such pain and discomfort in the right side that she was unable to perform her work."

slips of tongue, misdescription mistaken comparisons, and mistakes of language, hits other snags. For what is it he can't be *corrected* with regard to? It can't be his statement: it must be his intended statement or his belief. Now if he is at all specific he can still be mistaken; for he may believe that his sensation of sort g_0 is of sort g_1 , or that he had a sensation when in fact he had something else a bit like a sensation. The incorrigibility is only preserved by making his intended statement or his belief sufficiently *vague*. This can be done not only with sensation statements (e.g., "I had now a *something* or other"), but with other sorts of statements as well (e.g., "Something has happened to me now"). Incorrigibility and vagueness do not vary independently. It seems that as we increase incorrigibility we increase vagueness, and that in the limit an absolutely incorrigible "report" or "belief" is absolutely vague—not really a description, report, or belief at all. (v) A set of non-equivalent definitions of "*a*'s statement '*p*' is incorrigible," definitions based on criteria extracted from various authors, reveals that the exact sense of "incorrigible statement" lacks clarity. This point is argued in detail in the Appendix.

Other contrasts than public/private and corrigible/incorrigible contrasts, such as physical/mental and observable/theoretical contrasts may appear to block contingent identity hypotheses. They don't. If these predicates marked out properties which created difficulties for heterogeneous identifications similar difficulties would arise in the case of accepted scientific identities. The hypothesis (F_1) that temperature is mean kinetic energy does not run into difficulties over thermodynamical/mechanical and observable/theoretical contrasts: similarly (B)-hypotheses do not strike problems with these other contrasts. Subject-classifying and category predicates like "electrodynamical," "mental," and "subjective" do not specify properties which succeed in destroying contingent heterogeneous identifications.

On the contrary (β), if fulfilled, provides a condition under which a subject can be extended. Then objections based on classificatory predicates vanish because the classificatory properties cease to be exclusive. For example, with the advent of the reduction of temperature under the kinetic theory a phenomenon could correctly be classed as both thermodynamical and mechanical. Similarly if (α) and (β) are satisfied in the case of the identification of SENGHS and PROGS mental and physical are not exclusive. Indeed given (α) and

(β) objections based on classificatory predicates could be crushed simply by reclassification. Classifications are fluid: they change with contingent identifications, which destroy the exclusiveness of classifications. Nor are subjects as sharply separated as the objection would suppose. "Mental" and "physical" for instance do not mark out exclusive properties: various activities such as driving with care and reading are correctly classed as both mental and physical; and sensations are investigated in physiology. Next, subject classifications depend on the state of our knowledge and beliefs. Therefore classificatory properties which directly reflect this dependence are non-extensional. But contingent heterogeneous identifications are unaffected by variance of intensional or impure properties. For this reason too, observational/theoretical contrasts provide no basis for criticism of contingent identity claims: for the relevant properties are not extensional. Consider, to demonstrate this point, material equivalences between statements like "That is a violet light" and "That is electromagnetic radiation of wavelength approximately 400 Å." The predicates "is an observational statement" or "That . . . is observational" and "is a theoretical statement" do not transfer across the equivalence preserving truth-value. If, contrary to our view, one of these predicates were always to transfer across such material equivalences the observational/theoretical contrast would not prove an obstacle to contingent identifications.

Finally, although sensations generally have locations, e.g. the pain is in his foot, her head is giddy, a location problem still threatens (B)-hypotheses. For though a man's pain may be in his foot relevant physiological occurrences may occur all the way from his foot to pain centers in his brain. But if one location is less extensive than the other how can there be contingent identity? First, the *felt* location of a given pain is an intensional property of the pain, and so is discarded for contingent identity. According to *OED* ". . . is in —" in the relevant sense means "the part affected by . . . is —." This definition needs revision to allow for chiropractors' reports and for the man with an amputated leg who insists that the pain is in his foot. It seems that "*a*'s pain is in —" means "the part affected (the painful part) is felt (thought) by *a* to be —." The predicate "the part affected by . . . is felt by *a* to be *h*" demarcates an intensional property; therefore the synonymous predicate "*a*'s . . . is in *h*" also does. Second,

some physiological processes do appear to have felt locations as well as actual locations, e.g., those identified with pains at the extremities of the nerve fibers in which the processes occur; and *felt locations* could be suitably defined for some of those for which predication of a felt location would be non-significant.

IV

(B)-hypotheses are no more than contingent identities. Merely contingent identity statements are *contingent* statements. For if such a statement were not contingent but analytic, then the identity would hold for some modal properties: and then the identity would not be merely contingent. Since the restricted identity theory is merely contingent there is no reason why the terminology of the theory should conflict with rather than refine and complement everyday terminology for discoursing about sensations or the theory, if correct, with the truth of many everyday statements about sensations. Even if the identification were adopted as a criterion, e.g., of real sensations, it would not follow that everyday discourse must be abandoned and replaced by terminology drawn from the theory. The fact that in some circumstances we use occurrences of electrical discharges as a criterion for whether or not lightning occurs does not prevent or destroy phenomenal or ordinary descriptions of lightning, nor imply that we have to abandon such discourse and associated tests and speak instead in terms drawn from the theory. Similarly even if situations sometimes arose where we were to use physiological criteria to decide whether someone really felt a pain, these would not prevent us making the statements we do nor force us to withdraw or reject as false many of them. On the contrary, since "... occurs" and "... exists" specify extensional properties, it follows if identity (B) is true that SENGs occur and exist if and only if PROGs do. *Nor*, in general, does the introduction or establishment of contingent theoretical identifications *change the sense* of the observational *reducentia*. Just as there was, with "red," "hot," "lightning," though the senses of the words were determinate, room for further theoretical connections and identification, without change (or substantial change) of the sense of the words, so there is with sensation expressions room for heterogeneous identifications without change of sense of the expressions. In this respect an observational

expression ' x ' is *theoretically completable* without changing its sense, even though tests or criteria for x 's may shift once the theory becomes well-established and identity statements are normic. Introduction of normic connections, in place of merely contingent connections where the connections are well-established, is possible with cluster and multi-criterion words without change of sense of the words. Just this happens along with the development of theories. As in the lightning example, so with sensations there would be justification for introducing theoretical connections. There is not only room, but a need for them.

Subject to certain qualifications, e.g., relative to an embedding scientific theory P_1 given which the unspecified constants in representative sentences are specified, (B)-statements are *empirical*. A modified verification criterion can be used to establish this point. For example, (B)-statements are empirical because together with empirical statements about SENGs they (minimally) entail empirical statements about PROGs. (Compare tests for r -pains indicated in Sect. III.) But it does not follow, just because (B)-hypotheses are empirical, that there are not other hypotheses which are verificationally indistinguishable from them. Notwithstanding, it is frequently assumed that the *maximal empirical statement* which evidence could support can be formulated something like (J): Any sensation of sort $h(g_0)$ in x is correlated one-to-one with and is (approximately) simultaneous with a physiological occurrence of sort g_0 in x . That is, there is a class (presumably finite) of sensations of sort $h(g_0)$ in x , and a class of physiological occurrences of sort g_0 in x , and these classes are correlated by a one-to-one relation, say R_0 . Further, each item, PROG or SENG, occurs at approximately the same time as its correlate under R_0 , in the same time-slice Δt , say.

Difficulties infect the obtaining of a non-trivial correspondence because the occurrences involved are extensive, complex, and hard to distinguish from one another except in an arbitrary way, or, in the case of PROGs, in a way *just* based on patients' reports of SENGs. To highlight these difficulties consider the following. We assume first that at each time t , x_0 has a SENG there is some physiological occurrence in x_0 , second that x_0 only has one SENG at a time, and thus within bounds Δt containing t . Then we can set up a trivial correspondence satisfying (J) by defining as the correlate of SENG at t all physiological occurrences in x_0 in interval Δt . To avoid such difficulties the assumptions,

especially the second, must be dropped, much more detailed descriptions and specifications of SENGHS and PROGS and their patterns must be given, and the *occurrence and extent* of PROGS must, after initial calibration of instruments and determination of extent, be *marked off independently* of the occurrence and extent of SENGHS. Restriction to a single person x_0 should be dropped. Compare, as regards calibration and the setting up of correspondences, the beginnings of thermometry.

The thesis analogous to (J) for lightning runs: (H): Any lightning (flash) is correlated one-to-one with and is s.t. (i.e. spatio-temporally) coincident with an electric discharge (of sort) d (i.e., from an ionized cloud to earth, etc.). Here and below, "(s.)t. coincident" abbreviates "approximately (spatio-)temporally coincident." On the basis of (H) together with electromagnetic theory several facts about lightning can be explained and predicted. For example, given initial conditions whether the lightning will be sheet, forked, or chain, its striking point can be predicted. So given electromagnetic theory there are good reasons for advancing beyond (H). But to grant (H) on its own is by no means to grant (F_2): also, as required under (β), further features of lightning must be explained and predicted.

Traditional one-track mind-body theories, seen in miniature in the sensation case, can be reflected in the lightning example. For the following hypotheses might be put up on the strength of (H).

- (H₁) Between the *different* items lightning flashes and the occurrence of electrical discharges d there is a mere 1-1 correspondence and s.t. coincidence; and since respective correlates are different they are two.
- (H₂) Lightning is caused by (and s.t. coincident with) electrical discharges d (but not conversely).
- (H₃) Between (the different items) lightning and electrical discharges d there is two way interaction (and s.t. coincidence).
- (H₄) Lightning and electrical discharges d are just two (corresponding and s.t. coincident) aspects of the one underlying process.
- (H₅) (F_2), which implies that there is only *one* item.
- (H₆) Really there is no such (separate) thing as "lightning" but only electrical discharges d .
- (H₇) Lightning is nothing over and above electrical discharges d .

The bracketed expressions are included in an hypothesis just in case they are not rendered unnecessary by the remainder of the hypothesis.

Further hypotheses, e.g., single aspect, could be added.

A corresponding set of hypotheses (J_1)-(J_7) can be obtained for the sensations case (J), by replacing "lightning (flash)" and "electrical discharge d " in (H_1)-(H_7) by "SENGH" and "PROG" respectively. Comparing the two sets of hypotheses in detail is valuable. It is easy to envisage situations in which the dispute over (H)-(H_7) might be just as alive as disputes over (J)-(J_7), e.g., if our religion or cultural-history included lightning-worship (not science-worship) and people were still aesthetically interested in lightning.

Given suitable expansion or explication of such key expressions as "cause" in (H_2), "interaction" in (H_3), and "aspect" in (H_4) and given more explicit formulation of certain hypotheses, member statements of a set of hypotheses are verificationally equivalent, as (H_1) is verificationally equivalent to each member of the (H)-set. Roughly, statement S_1 *verificationally-implies* (*v-implies*) S_2 when S_2 has no observational consequences which S_1 does not have, i.e., when S_1 and S_2 entail inconsistent (direct) observational statements. Let $\mathbf{P}$ be a given consistent class of statements; e.g., the statements of theory P_1 . Then S_1 *v-implies* S_2 w.r.t. $\mathbf{P}$ (i.e., with respect to $\mathbf{P}$) iff all observational consequences of S_2 and $\mathbf{P}$ are observational consequences of S_1 and $\mathbf{P}'$, where $\mathbf{P}'$ is a modification of $\mathbf{P}$ obtained by at most uniformly replacing RST (equivalence) relations of a well-defined class of sentences of $\mathbf{P}$ by different RST relations (and adding analytic statements). S_1 is *v-equivalent* to S_2 w.r.t. $\mathbf{P}$ iff S_1 implies S_2 w.r.t. $\mathbf{P}$ where $\mathbf{P}'$ is the modification of $\mathbf{P}$ and S_2 *v-implies* S_1 w.r.t. $\mathbf{P}'$ where $\mathbf{P}$ is a suitable modification of $\mathbf{P}'$ i.e., if all observational consequences of S_2 and $\mathbf{P}$, are observational consequences of S_1 and $\mathbf{P}'$ and conversely.

There are two stages in establishing the required *v-equivalences*. In the first the (J)-hypotheses are shown to divide into two classes, the (J_1)-class and the (J_5)-class, such that each hypothesis of a class is *v-equivalent* to every hypothesis of its class. But there are indeterminacies in hypotheses of the (J_1)-class which can be resolved in two directions, either so that (J_1)-hypotheses differ verificationally from (J_5)-hypotheses or so that the hypotheses are *v-equivalent*. Thus in the second stage we replace (J_1)-hypotheses by (J_1^+)-hypotheses, and explain why (J_1^+) is *v-equivalent* (w.r.t. the embedding theory) to (J_5). To accomplish the first stage preliminary results are useful.

1. If S_1 entails S_2 then S_1 v -implies S_2 w.r.t. any $\mathbf{P}$. For if S_1 entails S_2 all consequences of S_2 are consequences of S_1 ; therefore all observational consequences of S_2 are observational consequences of S_1 .
2. If S_1 and S_2 are mutually entailing, or strictly equivalent, S_1 is v -equivalent to S_2 w.r.t. any $\mathbf{P}$.
The second part follows because differences between mutual entailment and strict equivalence show up at most in non-synthetic consequences, whereas all observational statements are synthetic.
3. If S_1 and S_2 are v -equivalent w.r.t. $\mathbf{P}$ and S_2 and S_3 are v -equivalent w.r.t. $\mathbf{P}'$ where $\mathbf{P}'$ is a suitable modification of $\mathbf{P}$ then S_1 and S_3 are v -equivalent w.r.t. $\mathbf{P}''$ where $\mathbf{P}''$ is a suitable modification of $\mathbf{P}$ (or $\mathbf{P}'$).

Because of 3 it suffices to establish a chain of v -equivalences.

To sharpen (J₁) we define mere correspondence: classes w and v *merely correspond* iff they correspond one-to-one, i.e., there is a one-to-one relation between them, but there are elements x of w and y of v such that for some extensional property f , ' $f(x)$ ' and ' $f(y)$ ' are significant but x has f and y lacks f ; then:

If w merely corresponds to v , w is not contingently identical with v , i.e., $w \not\equiv v$.

In the *sketch arguments for the first stage* $\mathbf{P}$ may be any set of statements consistent with the conclusions of Sect. I. Now x and y *interact* iff there are causal relations between x and y . Since then (J₂) entails (J₃), (J₂) v -implies (J₃). In fact hypotheses (J₂) and (J₃) differ at most as to the direction of the causal relation in some instances. But in the case of sensations, causal relations are from physiological processes to sensations (see Sect. I). Therefore (J₂) and (J₃) do not differ verificationally here, i.e., differences between epiphenomenalist and interactionist hypotheses only appear in larger contexts. In coping with (J₂) it simplifies matters and seems permissible to use a regularity analysis of "cause": roughly, where determinate causal relations are not observable v causes w iff there is a one-to-one (lawlike) relation R_0 correlating v with w such that if xR_0y then x does not succeed y . (J₂) splits into two cases. Either $w \cong v$ and (J₂) coincides with (J₅); or $w \not\equiv v$ and (J₂) is v -equivalent to (J₁). For then (J₂) entails (J₁). (Alternatively use the distinguishing property "... causes y_1 "). Conversely to show (J₁) v -implies (J₂) use the fact that if y_1 is the correlate of x_1 under R_0 then if x_1 is t.coincident with y_1 then x_1 does not succeed y_1 , and use a distinguishing property. (J₄) also splits

into two cases. Either the aspects are identical with one another and so with the underlying process, and (J₄) coincides with (J₅); or the aspects differ, and (J₄) is v -equivalent to (J₁). For use:

$$R^2 \in R. R = \check{R} \equiv . (\exists S) . S \in Cls \rightarrow 1 . R = S/\check{S}.^{12}$$

If (J¹) holds, there is a transitive, symmetric relation R_0 relating w and v , and therefore there is a many-one relation S relating w and v to some underlying item. Thus, short of further explanation of "aspect" to exclude the equation, w and v are S -aspects of an underlying item, i.e., (J₄) holds. Conversely if (J₄) holds w and v are related to an underlying process by aspect relations, which we can combine as a many-one relation S_0 . Then $R_0 = S_0/\check{S}_0$ is a relation having the required properties for (J₁). Alternatively consider the relation product R_1 of aspect relations S_0 and T_0 , and make use of the fact that the aspects are corresponding, t.coincident, and different. (J₇) implies (J₅). If x 's are nothing over and above y 's therefore x 's have no properties that y 's do not. All properties that y 's have in addition to these properties are properties like f whose predicates are not significantly predicable of *reducentes*. For otherwise not- f would be a property of x 's but not of y 's, contrary to hypothesis (J₇). Therefore x 's and y 's have all extensional properties in common; and $w \cong v$. The converse is clear, provided by "nothing" is meant "no extensional properties." In (J₆) the qualifying word "separate" is crucial; otherwise (J₆) could be refuted given (J₅). Consequently also (J₆) should not be formulated "... do not exist; only—exist"; a qualification like "independently" is needed in the first clause if v -equivalence is to be preserved. (J₅) implies (J₆); for if x and y are identical there are no extensional features separating them; therefore they are not separate. Conversely that x and y are not separate implies that they are coincident.

Two classes of hypotheses emerge, the one-item (J₅)-class and the two-item (J₁)-class. Because the formulation of (J₁) is insufficiently explicit, (J₁) and (J₅) are not clearly v -equivalent. V -equivalences of two-item hypotheses only follow because of similar inexplicitness in other hypotheses of the class. The salient point is that it is not made explicit whether or not x_1 displays a similar observational pattern to its correlate y_1 . Similarity has yet to be explained; but suppose, to illustrate, we discovered that in a time interval of occurrence

¹² A variant of 72.66 of A. N. Whitehead and B. Russell, *Principia Mathematica*, 2nd ed. (Cambridge, Cambridge University Press 1925). The symbolism for relations is drawn from this text.

t_1 to t_{1+1} a typical PROG (or y_1) had a quite different intensity graph from its correlated SENGH (or x_1), e.g., the positive cycle of a sine wave opposed to a saw wave. Such data might be sufficient to falsify (J_5) -hypotheses and/or to require modification of underpinning theory P_1 . But the data would not falsify or disconfirm (J_1) -hypotheses unless these hypotheses also implied that the observational patterns of correlates were isomorphic. To progress we shall have to sharpen (J_1) -hypotheses. Either (a) (J_1) -specifications should be such as to require that correlates are observationally isomorphic, or (b) such as to require that they are not always observationally isomorphic. In case (b), (J_1) is not v -equivalent to (J_5) because (J_1) together with the relevant theory implies that there are correlates which have differing observational features, whereas (J_5) implies that there are not. Those who have espoused (J_1) -hypotheses do not seem prepared to commit themselves to (b); nor should they, unless they think that there is mainly an empirical difference between (J_1) and (J_5) , and that there is some evidence in favor of (b). So let us recast (J_1) -hypotheses so that they conform to (a). We abbreviate "observational" by OBS, and "observationally isomorphic" by *o-isomorphic*. Then we introduce a definition designed to capture essentially what we mean by "having identical observational features":

x is *o-isomorphic* with $y \equiv_{\text{DF}}$
 $(\forall \text{ EXT \& OBS } f). [\Sigma^* f(x)^* \& \Sigma^* f(y)^*] \supset [f(x) \equiv f(y)].$

Next we replace (J_1) – (J_4) by (J_1^+) – (J_4^+) , where (J^+) is obtained from (J^-) by adding to the formulation of (J^-) :

" $\& (\forall y, x)(yR_0x) \supset [y \text{ is } o\text{-isomorphic with } x].$ "

In arguing for the v -equivalence of (J_1^+) and (J_5) we depend on the point that (J_1^+) and (J_5) are not themselves purely observational statements. Otherwise, in virtue of preceding definitions, they could not be v -equivalent. It follows that (J_1^+) and (J_5) differ (extensionally) only with respect to non-observational properties.

The relevant differences between (J_1^+) and (J_5) will appear in their embedding theories. Let the complete embedding theory of (J_5) be P_5 , where P_5 includes theory of physiological processes P_1 and requisite logico-mathematical and auxiliary theories. Observational consequences of $P_5 + (J_5)$

are obtained by making use of initial observation conditions. The initial conditions will be the same irrespective of whether (J_1^+) or (J_5) is used, but the theories used to derive observational consequences will differ slightly according as (J_1^+) or (J_5) is used. We want to show that (J_1^+) is v -equivalent to (J_5) with respect to P_5 . Thus we need to show that (J_5) can be replaced by (J_1^+) and the embedding theory by P_5' without diminishing the class of observational consequences, and similarly under replacement of (J_1^+) and (J_5) w.r.t. P_5 . Like reduction sentences (J_1^+) and (J_5) are only needed at the beginning and end of any derivation of observational consequences because other mathematical transformations can be carried out within the respective embedding theories P_5' and P_5 without making use of *reducendes* expressions *at all*. Because of results like: $w \text{ sm } v \& v = u \supset . w \text{ sm } u$, where 'sm' abbreviates "is similar to," correspondences will be preserved. Reformulate derivations of observational consequences so that all applications of (J_1^+) , or of (J_5) , occur at the beginning or end of respective derivations. There will be at most a finite number of applications of (J_1^+) , or (J_5) , at the beginning and end of any such derivation. Such transformations of proofs can be carried out effectively; thus a principle of substitution of identicals for *reducenda* with *reducendes*, the main technical advantage of identity relations over other equivalence relations, is not required.¹³ Suppose, then, in a derivation of an observational consequence using $(J_1^+) + P_5$ we have rearranged the derivation so that (J_1^+) occurs at only initial and/or final stages, and all intermediate steps are carried out without using *reducenda* expressions. We may in rearranging have to make use of simple equivalences from the theory of relations. Next, systematically replace one-to-one correspondence relations appearing in (J_1^+) connections by contingent identity relations, and thus (J_1^+) by (J_5) and P_5' by P_5 . Derivations of observational consequences are not destroyed, for a derivation could only be altered at initial or final stages if a property serving to distinguish mere correspondence from identity appeared essentially in the correlation. This could not happen at the end since the correlation is here employed in reaching observational consequences; so that there would have to be observational properties with respect to which (J_1^+) differed from (J_5) , contrary to our previous finding.

¹³ In fact a principle of substitution of identicals is all that is needed in first order functional calculi with identity to distinguish identity from other equivalence relations; and in the simple calculus of equality identity and equivalence relations coincide. See, e.g., A. Church, *Introduction to Mathematical Logic*, vol. 1 (Princeton, Princeton University Press, 1956), p. 281.

And the derivation could only be upset at the beginning if a distinguishing property were essential in getting from initial observational statements, or in reaching observational consequences. But then it would be an observational property contrary both to the requirement of *o*-isomorphism and to the neutrality of initial conditions. Similarly starting with derivations using $P_5 + (J_5)$ replacement of contingent identity relations in (J_5) connections by one-to-one correspondences preserving *o*-isomorphism, and thus replacement of (J_5) by (J_1^+) and P_5 by P_5' , does not diminish the class of observational consequences. This completes the sketch argument for *v*-equivalence.¹⁴

It is not true that correspondence hypotheses like (J) or (J_1) make fewest assumptions and approximate best to empirical data; and that therefore such hypotheses are all we are warranted in opting for. Unless revised as in (J_1^+) correspondence hypotheses may not even take account of all observational data. With hypotheses which are *v*-equivalent w.r.t. the requisite theory there is no basis for claiming that what has been established scientifically just is a particular one of them, e.g., a correspondence hypothesis, or that such a one has privileged status as "closest to the scientific facts." Such *v*-equivalent hypotheses have equally good empirical standing. The fact is that the (putative) scientific findings have *here*, in (H^+) and (J^+) , been presented in a way which, though fairly neutral (there are limitations on the inter-theory neutrality that can be got in the formulation of scientific hypotheses containing *theoretical* [not observational] expressions or expressions with important theoretical connections), is more favorable to some hypotheses than others, as to a correspondence than to an identity or double aspect thesis. For they suggest that there is mere correspondence and thus that there are *two* things which are correlated. Since this holds equally for

(H) -hypotheses as for $(J)^-$, it is not true that to grant (H^+) is just to grant (H_5) . To grant a statement is not thereby to grant any statement *v*-equivalent to it w.r.t. accepted scientific theories, or there could hardly be so much philosophical rivalry. These factors should also emphasize difficulties for those who adhere to hypotheses of each set which do not correlate numerically, for example, to both (H_5) and (J_1) .

On what bases do we *choose* between (H_1^+) – (H_7) or (J_1^+) – (J_7) ? How do we decide what to *say* in (J) -cases where usage is not determinate or well-founded. Short of introducing, as with (α) and (β) , sharper criteria, e.g., for whether *one* process or *two* processes occur, criteria which when accepted settle the issue, there is no satisfactory way of resolving the issue apart from the weighing of cumulative reasons for deciding in one way or on one hypothesis rather than another. And the defense of the introduction of such sharper criteria comes down to the weighing of such cumulative reasons. The issues are characteristic *conflict* or decision *issues* in which the outcome, if any, is not fixed by empirical matters to more than *v*-equivalence though empirical considerations may be important in reaching a reasoned verdict. There is a relevant difference between the (H) and (J) cases. The (H) -group at present provides us with an over-determined conflict case, i.e., the case has normally been resolved in favor of (H_5) and there is a fairly well-established usage. But the (J) -group presents us with an under-determined conflict case, where there are rival usages linked with different criteria and comparisons. This is not the only difference between the groups; for token-sensations are private and relatively incorrigible.

Serious conflict issues are not satisfactorily resolved either by chance (e.g., tossing a coin) or by applying Occam's razor or crude parsimony principles.¹⁵ To appreciate how more balanced

¹⁴ If *v*-equivalence were so defined as to require only sameness of observational consequences, Craig's theorem could be used to show *v*-equivalences.

¹⁵ Since we have been treated to large doses of these applications in recent literature we present briefly some defects in the views that one does, ought to, or must choose between hypotheses using Occam's razor (abbreviated *OR*). (a) *OR* does not discriminate between (H_4) and (H_5) and (H_7) , if aspects or appearances are not entities. (b) *OR* is not a panacea. If applied indiscriminately it could effect all sorts of reductions many of which have little plausibility. In order for it to be of any use it has to be heavily supplemented by further conditions; in particular, criteria have to be supplied for what is to count as an *entity*, on which issue *OR* usually begs the question, and for *necessity*. It suggests a pragmatic criterion for necessity, which is objectionable since the existence of entities is not a matter of their usefulness or practical necessity. (c) Unless these conditions are supplied *OR* looks remarkably like an unwarranted stipulation. We can ask for, but not seriously expect, a justification for *OR*. And there is nothing to stop us from setting up a rival principle. (d) *OR* is much too nominalistically oriented. For strictly we cannot either multiply or reduce most non-abstract entities or items; we can only as a rule multiply or reduce our terminology for talking about them. (e) *OR* is unnecessary. We can settle conflict issues, and handle "hypotheses" *OR* is often invoked to eliminate, without appealing to *OR*. So if a principle is an entity *OR* includes itself in its own scope. For related reasons parsimony principles have dubious value in resolving conflict issues unless carefully hedged.

decisions between rival hypotheses may be reached consider that major issue which turns on whether to say "one thing" or "two things," "one process" or "two processes." Deciding for identity of items as in (J_5) implies but is not tantamount to deciding in favor of "one item"; for if x_0 and y_0 are identical (not two things with identical properties apart from spatial or temporal properties) then x_0 and y_0 are one. If x_0 and y_0 only correspond then x_0 and y_0 do not have all properties in common, and are therefore distinct. Thus adoption of (J_1^+) or (H_1^+) implies deciding on "two distinct occurrences."

To appreciate reasons for saying "one occurrence" rather than "two occurrences" consider a series of cases. Consider first a furniture remover and his reasons for saying that we have two pictures but only one table, not both an Eddington table and a Stebbing table.¹⁶ Consider next the reactions of a lightning-remover expert to various of hypotheses (H^+)–(H_7). His client, an adherent of one of (H^+)–(H_3^+), says: "All right, that thing you put up gets rid of the electrical discharge and the crackle on the radio; but what are you going to do about the lightning and its burning effects? Don't we need some protection against that too?" In these circumstances we tend to seize on (H_4)–(H_7). None of (H^+)–(H_3^+) would normally be sufficient to allay the client's anxiety. We should have to explain somehow how we have complete control over the lightning once we have control over electrical discharges and that rendering one of them harmless renders the other ineffectual. Even a causal hypothesis like (H_2^+) doesn't surmount these difficulties at all easily. Connections between correlated items of (H^+)–(H_3^+) remain obscure, even mysterious, and insufficiently elucidated. Briefly, (H^+)–(H_3^+) are *explanatorily inadequate*. Now consider the case of someone who has a pathological fear of after-images, has convulsions when they occur, etc. We tell him about his retinal structure, optic nerve, and visual cortex and instruct him not to stare at bright objects for any length of time. He says, "That's all very well, but it's not my retina and whatnot I'm worried about—it's those weird visions I keep getting." Or consider an exponent of one of (J^+)–(J_3^+) who gets severe pains in his arms. His doctor gives him a pain-killer (-remover) like morphine and a rudimentary physiological explanation. But the patient says, "I understand how this stuff goes into my blood-stream, how it affects my neurons, but I'm

still at a loss to know how it will stop that dreadful stabbing feeling in my arm." How could we help him using his hypotheses? These hypotheses are unsatisfactory here because they make it *look* as if there is something further to be *explained*. They encourage questions like: *Why* and *how* does this correspondence occur? We can't answer this question without either blocking it with, e.g., "the correspondence is ultimate (unexplainable)" or "there are fundamental category differences (which are unexplained, unexplainable, or don't need explanation)," or else appealing to some further identity and connections. The first alternatives each create serious problems: some of them violate the principle (of sufficient reason) that every phenomenon has a (rational) explanation, and so isolate themselves from support; most give no account or explanation of the requisite ultimacy or fundamentality, and so on. With contingent identity these problems are avoided, since if (α) and (β) are satisfied, explanatory relations are coupled with identity relations. Since, however, the relevant explanations are mostly causal explanations (c.f., Sect. I, 8) it is imperative to allow, as we have, that " x causes y " is *consistent* with " $x \cong y$." Because it is quite fashionable to so characterize "cause" and "explanation" that explanatory relations are inconsistent with identity relations it needs stressing that such a course has devastating consequences; in particular that the making of heterogeneous identifications in science is inconsistent with the use of these same connections (in the way they regularly are used) to provide explanations, and that a full materialism like (M_1) is inconsistent with much of psychology.

Many other *reasons tell against* certain (H)- and (J)-hypotheses. Correspondence hypotheses, (H_1^+) & (J_1^+), as well as being explanatorily unsatisfactory, appear mysterious. Any mere correspondence can be destroyed without destroying the correlates; but an identity cannot significantly be destroyed, only the item can be. If a correspondence can be destroyed it makes sense to ask what keeps it going, what explains it, questions which are not significantly asked of an identity. *This* is why an identity succeeds in blocking requests for further explanation. For every equality and correspondence there is an underlying identity. For if v and w are correlated by one-to-one relation R_1 then correlated items are identical apart from certain extensional features which one has and the other lacks. If (J_1^+) were correct they would not be

¹⁶ See L. S. Stebbing, *Philosophy and the Physicists* (London, Methuen, 1937); and J. Wisdom, *op. cit.*, p. 67.

observational properties. Suppose (J_1^+) is correct, then v and w are identical in all extensional observational respects. Consider the classes v' and w' consisting of items exactly like those of v and w except that the distinguishing theoretical features are not truly predicable of them. Then $[v' \cong w']$ and v' is o -isomorphic with v , and w' with w . It is this *identity* which accounts for the correspondence and o -isomorphism of SENGs and PROGS. Furthermore, there are difficulties in setting up a correspondence in ways which do not seem to rely on identifications. For we decide, e.g., where physiological processes end, initially on the strength of when patients say their sensations finish, etc. Here it looks as if we have made a tentative identification. With (H_2^+) and (J_2^+) , either unsatisfactory regularity-analyses of causation are offered, in which event the hypotheses face difficulties analogous to those set out with respect to (H_1^+) and (J_1^+) , or else "cause" is characterized as a highly determinable word and central examples of causation are explained in terms of observationally related items or in terms of a story linking such items. But here we encounter a serious difficulty, namely, that without an identification like (H_5) there remain *gaps* in the observations and the story which cannot be bridged. To complete the causal story and to bridge the gaps an identity hypothesis has to be introduced. In initial formulation (H_3^+) and (J_3) suffer from shortcomings which mar most of the hypotheses, namely, obscurity. If interaction is causal interaction then (H_3^+) and (J_3^+) are confronted by the same difficulties as in (H_2^+) and (J_2^+) . It is not enough to rely on mechanical and electrical analogies like those provided by automatic control and servo-mechanisms such as thermostats and feedback circuits where the requisite explanatory details can generally be filled out. Among serious defects of (H_4^+) and (J_4^+) are that they provide no details at all of the underlying processes or of how the aspects are aspects of these processes. The analogies on which double aspect hypotheses are based, e.g., ambiguous representations, do not satisfactorily transfer to (H) and (J) cases. What is the single ambiguous item, like the drawings of the Maltese cross or duck-rabbit or the cloud-patterns, which have various aspects or interpretations? Of what are sensations appearances or aspects? In normal senses of "aspect," sensations are not aspects; so we should have to introduce a new sense of "aspect." Such a new sense can be used to advantage to collapse (J_4^+) into (J_5) . (H_6) and (J_6) have

to be carefully formulated if they are not to conflict with everyday observations. When carefully presented they are very close to (H_5) and (J_5) ; still they tend to sacrifice a main advantage of identifications: that of providing explanations, while not looking like eliminations, of familiar phenomena. Differences between various hypotheses grouped together under (J_6) , e.g., sophisticated stimulus-response and behaviorist hypotheses, only emerge under more extensive reductions of mental items. Because vague, since "over and above" can be expanded in different ways, (H_7) and (J_7) carry misleading suggestions, e.g., they suggest reductions of sense, they look like old-time analyses. If, for instance, (J_7) were replaced by "SENGs have no properties over and above those of PROGS" it would be wrong. It is correct at best when the relevant properties are restricted to extensional properties.

Accumulative reasons weigh out in favor of (H_5) and (J_5) . Therefore, as is usual, we adopt (H_5) ; and if the requisite theory P_1 were available and had some confirmation we should definitely select (J_5) .

Accepting (J_5) we can effect a *synthesis of hypotheses*. Since the intensional and significance features of SENGs or PROGS provide extra irreducible features possessed by the one but not by the other, we can combine (J_5) with a familiar double aspect hypothesis. The extra features which are specified by intensional sentence frames are irreducible because the expressions referring to each item of the contingent identity (J_5) do not have the same sense, and therefore cannot be substituted for one another in all intensional sentence frames preserving their sense. (J_5) implies this double aspect hypothesis. But the identity hypothesis is compatible with double aspect hypotheses only if the additional irreducible features are intensional or significance features.

(J_5) implies a correspondence hypothesis if "correspondence" is so redefined that v and w only correspond if they coincide but are not strictly or regularly identical. A causal hypothesis according to which PROGS cause SENGs is implied by (J_5) because (J_5) guarantees that (β) is satisfied and the requisite explanations under (β) are causal. And so on for the synthesis of other v -equivalent hypotheses with (J_5) . Each of these hypotheses, then, has picked out just some features of the contingent identity. Unless they insist (as they usually do) that the aspects they select are the only features they are all reconcilable with the contingent identity hypothesis (J_5) .

APPENDIX

Symbolize “*a*’s statement ‘*p*’ is incorrigible” by “ $I_a[p]$,” and the *k*th definition of “ $I_a[p]$,” by “ $I(k)_a[p]$,” “*a* (genuinely) believes *p*” by “ $B_a[p]$,” “*a* is mistaken about (the truth of) *p*” by “ $M_a[p]$ ” and “*a* knows that *p*” by “ $K_a[p]$.” If *a*’s statement that *p* is to be incorrigible then lying, deceiving, etc., must be excluded, and *a*’s report must be sincere; therefore we want at least to have $[I_a[p] \gg B_a[p]]$.¹⁷ Now compare definitions reached:

$I(1)_a[p]$ FOR $B_a[p] \& (B_a[p] \gg K_a[p])$

$I(2)_a[p]$ FOR $B_a[p] \& (B_a[p] = [p])$

$I(3)_a[p]$ FOR $B_a[p] \& (p \gg K_a[p])$

$I(4)_a[p]$ FOR $B_a[p] \& \sim \Diamond M_a[p]$

$I(5)_a[p]$ FOR $B_a[p] \& (B_a[p] \gg K_a[p]) \& ([p] \gg B_a[p])$

Since $[K_a[p] \gg [p]]$ and $K_a[p] \gg B_a[p]$ follow from the definition of “ $K_a[p]$ ” and $[B_a[p] \supset [M_a[p] = \sim p]]$ from the definition of “ $M_a[p]$,” it follows that $[K_a[p] \gg \sim M_a[p]]$ and $[I(4)_a[p] \supset \Box[p]]$. Thus $I_a(4)$ “disastrously” excludes synthetic incorrigible statements. $I(5)$ implies $I(1)$, $I(2)$, and $I(3)$. Since $[I(1)_a[p] = .B_a[p] \& (B_a[p] \gg K_a[p])]$ and $[I(3)_a[p] = .B_a[p] \& ([p] = K_a[p])]$ any two of $I(1)$, $I(2)$, and $I(3)$ imply the third. But $I(1)$, $I(2)$, $I(3)$, $I(5)$ are not equivalent unless knowledge is erroneously equated with true belief, i.e., unless $[K_a[p] = [p] \& B_a[p]]$. $I(2)$ and $I(3)$ are defective in that it doesn’t follow from them that *a* knows *p*, i.e., we do not have for $I(2)$ and $I(3)$ $[I_a[p] \supset K_a[p]]$. Since $I(5)$ implies $I(1)$ consider $I(1)$. It has the desired consequences $[I_a[p] \supset K_a[p]]$, $[I_a[p] \supset [p]]$ and $[I_a[p] \supset \sim M_a[p]]$; but it also gives

(9): $I_a[p] \supset \sim \Diamond [B_a[p] \& [\sim p]]$.

Opposing (9) is the very plausible

(10): $\Diamond [\sim p] \supset \Diamond [B_a[p] \& [\sim p]]$; roughly, it is

always logically possible to believe what is false. If (10) is true incorrigible synthetic statements are annihilated. Using $[\Delta[B_a[p]]]$, $[\Diamond [\sim p] \supset [\Diamond B_a[p] \& \sim p]]$ follows; but to prove (10) some thesis equivalent in strength to it is required. Given (10) a weakened version of: if *a* can’t be wrong he can’t be right either, viz.

(11): $[\Diamond \sim p] \supset [\sim \Diamond [B_a[p] \& \sim p] \supset \sim \Diamond [B_a[p] \& p]]$, follows. (11) is also sufficient to eliminate incorrigible synthetic statements. For (11) is equivalent to $[[\Diamond \sim p] \supset [[B_a[p] \gg [p]] \supset [B_a[p] \gg [\sim p]]]]$ so that it follows $[[I_a[p] \& \Delta[p]] \supset [p \& \sim p]]$; therefore $[\sim \Diamond [I_a[p] \& \Delta[p]]]$. $I(5)$ is less satisfactory than $I(1)$ since $[I_a[p]]$ implies $[M_a[p] = \sim B_a[p]]$ and $[B_a[p] = \sim \Diamond M_a[p]]$ (compare (iv)). Finally these definitions of $I_a[p]$ make it plain that incorrigibility is not an extensional property.

D. M. Armstrong has proposed in comment a contextual definition of ‘ $I_a[p]$ ’ which can be symbolized:

$B_a[p] \supset [I_a[p] \equiv .B_a[p] \gg [p]]$. But since it is equivalent, using $[I_a[p] \gg B_a[p]]$, to $[I_a[p] = .B_a[p] \& (B_a[p] \gg [p])]$ it fails to guarantee $[I_a[p] \supset K_a[p]]$. Also (9) and its consequences ensue.

Because of the consequences of (10) and (11), incorrigibilists will tend to reject both them and the arguments which support them, e.g., the Wisdom-Wittgenstein defense of (11) and the brief argument presented by Smart and others for what amounts to (10). As these defenses are commonly presented this is not a difficult task; the Smart argument, for instance, is so formulated as to rely on Hume’s separability principle (Hume’s separator), a principle which is valid at best under restricted conditions. (See A. Pap’s criticism of the separator in *Semantics and Necessary Truth* [New Haven, 1958], pp. 190–191.) All this is not to deny that (10) and (11) do have defenses of some plausibility.

University of Sydney

and

University of New England

¹⁷ Throughout this Appendix, ‘ $\gg$ ’ is used for strict implication and ‘ $\equiv$ ’ for strict equivalence.