

IDENTIFYING MENTAL STATES: A CELEBRATED HYPOTHESIS REFUTED

Irwin Goldstein

Today *functionalism* commands the attention behaviourism and the mind-brain identity theory did in the 1950s and 1960s. It is the most respected position on the mind-body problem.

A portrayal of what determines a mental state's identity is central to the collection of ideas that make up functionalism. Most functionalists think an event's causes and effects, its 'causal role', determine whether it is a mental state and, if so, which kind. Proponents say an event is a 'twinge', 'belief', or any other kind of mental state solely by reason of its bodily causes and effects (or 'inputs' and 'outputs').

Analytic functionalists think a sentence that specifies a mental state's identifying causes and effects is analytic. On this view a psychological word has a sense ('descriptive meaning') that determines the word's reference. The causes and effects definitive of a headache *qua* 'headache' are knowable *a priori*, simply by knowing what 'headache' means.

The analyses Saul Kripke and Hilary Putnam give water, gold, and other 'natural kinds' guide *empirical functionalists*. Kripke assigns a natural kind a 'species essence'. A natural kind's essence is knowable only *a posteriori*. Science suggests water's essence is H₂O. No properties essential to a natural kind are knowable *a priori*. A natural kind word has no sense to determine its reference.

Empirical functionalists think experiences are Kripkean natural kinds. A psychological word has no sense to determine its reference. An experience has an essence we can identify only *a posteriori*. 'The essence of the mental — like the essence of a natural kind such as iron — cannot be discovered *a priori*', writes Henry Jacoby [9, p.273]. Empirical functionalists think an experience's causes and effects form its empirically discovered essence.

Functionalists believe psychological words are substance neutral ('topic neutral'). An experience is a 'pain' or a 'thought' in virtue of its causes and effects, not its intrinsic character. We do not assign an experience intrinsic properties when we specify it with a psychological word. Functionalists conclude people do not commit themselves to the falsity of orthodox materialism when using everyday psychological words to specify their experiences.

Most functionalists are orthodox materialists. They think that what has a mental state's causes and effects in sentient beings we know, is a neural event. In other creatures something else, even an immaterial substance, might have these causes and effects and so *be* an experience.

The *orthodox* materialist embraces the neuroscientist's ontology. The sorts of

physical properties neuroscientists routinely assign brain events, the orthodox materialist recognizes. What might interact with neural processes in the central nervous system, according to the types of principles neuroscientists recognize, he recognizes. Functionalists think we can specify the properties constitutive of a pain, thought, or any other mental state by citing its causes, effects, and orthodox neural properties.

I intend to refute the functionalist's causal principle and to formulate and defend an alternative view of how we specify experiences. Analytic or empirical, the functionalist's principle is incorrect. Mental states *have* causes and effects. (When people use 'Mental states are functional states' to say only this, they say something trite.) However, a kind of experience's defining or essential properties do not lie solely in its causes and effects.

We identify neither *sensations* nor *thoughts* and other cognitive states exclusively by their causes and effects.

I confront and refute the functionalist's orthodox materialism. Properties essential to some experiences are not orthodox material properties.

I. An Experience has A-causal Properties

People sometimes speak as though a kind of experience might have no properties other than its being an effect of some events and a cause of others.¹ They think a person might specify all of an its immediate properties by specifying its causes and effects. Any suggestion that an experience has *other properties*, too — properties in addition to its causes and effects — they resist.

Whatever acts as a cause or an effect must have some *interior* — a nonrelational, self-contained, intrinsic nature. It must have 'a-causal properties', properties other than its causes and effects.

Causation is a *relation* between two distinct events or objects.

Suppose T has the relational property of being taller than S. It cannot be that T has no properties other than this relational property, 'being taller than S'. A causal relation is similar. When C causes E, it cannot be that C is comprised of this single property, 'being the cause of E'. C and E must be distinct events. C must be complete in itself: It must have a nature that enables it, in principle, to exist outside this relation with its intrinsic properties intact.

There is more to a pain, thought, or visual experience than its causes and effects. An experience has an interior. It has intrinsic, nonrelational properties. It has a-causal properties. A failure to affirm these truisms — unequivocally — nourishes the debate over functionalism.

How could people overlook these truisms? Here are three reasons.

¹ 'What can physics tell us about (electricity)? It can tell us what causes it . . . It can tell us about its effects . . . But it can no more define electricity than psychology can define kinds of experience. All these two sciences can say is that electricity and experiences play such-and-such roles in the laws which, among other things, link them to their causes and effects. That is all they can say; and that, I think, is all there is to say. There is no such extra thing as what electricity, or a kind of experience, is . . .', writes D.H. Mellor [15, p.14].

First, people routinely portray functionalists as defining *pain*, *thought*, and other *experiences* rather than *the words* 'pain' and 'thought', and other *words* that specify experiences. People assume a functionalist presents what he thinks is a complete definition of pain. People regard a complete definition of pain as an exhaustive description of pain.

Even if 'thought' *did* specify an event exclusively by its causes and effects, we would not exhaustively describe thought by exhaustively defining the word 'thought'. Thought would still have a-causal properties. An exhaustive definition of a *word* is very different from an exhaustive description of the kind of object we refer to with the word. A person who exhaustively defines the *word* 'war' has not thereby exhaustively described war. A person does not exhaustively describe *pain* by presenting an exhaustive definition of *word* 'pain'.

Second, some people blend 'We identify an experience by its causal role' and, what is very different, 'An experience is a causal role'. The latter borders 'there is no more to an experience than its causal role'.

A toothache is not some set of causes and effects or 'a causal role'. It is an introspectable event present for some protracted duration. This event *has* causes (such as tooth decay) and consequences (among them, visiting a dentist). When I say tooth decay caused Sheena's toothache, I refer to a specific introspectable event and say tooth decay caused it. This event has an interior with a-causal properties.

Third, some materialists view the causal principle not only as support for orthodox materialism but, for various reasons, a *requirement* for it. Threats to the causal principle they handle as threats to orthodox materialism. An admission that an experience has interior a-causal properties they regard as a threat to orthodox materialism. They fear critics of materialism will claim these residual, intrinsic properties are immaterial. Or irreducibly psychical.

These materialists reason from a mistaken premise. The admission that an experience has interior a-causal properties is not itself a threat to orthodox materialism. What has interior a-causal properties is not *ipso facto* immaterial or irreducibly psychical. Indeed, were thought a neural event, it would have interior a-causal properties. Neural events have them.

II. Experiences Divide Into Discrete Kinds in Many Ways

Some functionalists believe experiences divide into discrete kinds solely according to their causes and effects. These people cannot be right. Insight into how a class of objects divides into distinct kinds of those objects enables us to recognize this.

Cars, injuries, beliefs, sensations, and other types of objects divide into species, or kinds, in many ways. For every kind of property cars have there is a distinguishable 'kind of car'. We identify 'kinds of cars' when we talk of fast cars, red cars, small cars, luxury cars.

Experiences divide into discrete kinds in many ways. For every kind of property experiences have there is a distinguishable kind of experience. Experiences have various kinds of a-causal properties. Hence, there are kinds of experiences that could be identified by their a-causal properties.

III. 'We have no names for experiences identified by a-causal properties'

A functionalist may concede that we could identify sensations, thoughts, and other experiences by their a-causal properties. He may claim merely that we do not identify experiences this way with existing psychological words. Michael Tye [26, p.94] says functionalists direct their thesis at 'our everyday psychological vocabulary'. Understood so, functionalists claim that *with a particular set of words* — 'our everyday psychological vocabulary' — we specify experiences exclusively by their causes and effects.

This is an odd thesis. Which vocabulary is 'ours'? Is it English? What should French philosophers think of functionalism? Does each functionalist address languages *he* knows? Do functionalists extend their principle to all languages? How many do they know?

Anyway, as a claim about English or any other language, the causal principle is not self-evident. Many people regard the thesis as deeply counterintuitive.

We *could* pick out thoughts, intentions, and other mental states by their a-causal properties. There is no proof that we do not. Proponents back the causal principle with little *positive* argument. (Causal theorists defend their thesis by fielding objections.) Proponents do not appeal to facts more basic than the causal thesis and deduce it from these.² That causal analyses avoid some objections that refute behaviourism hardly proves the causal theorist's principle. Parallelism avoids some objections that attend interactionism.

David Lewis candidly admits he does not prove his causal principle [12, p.23]. He calls it a 'working hypothesis' [13, p.213]. The principle is just that. It is a hypothesis that is neither self-evident nor backed by substantial positive argument.

IV. An Experience has Introspectable A-causal Properties

Lewis thinks an experience's intrinsic character is a neural event's intrinsic character [14, p.506]. Many would agree. However, we do not ordinarily isolate an experience directly by a neural event's interior properties. People with no conception of neural states identify their experiences. Indeed, David Armstrong notes that 'in introspection we are not aware of mental phenomena as material states' [1, p.158].

Sensations and other mental states have an *introspectable interior*. A person has direct, introspective access to his sensations, thoughts, emotions, and intentions.

Most of the properties by which we identify experiences are introspectable. When a child reports a stomach-ache, he identifies his sensation by its introspectable, stomach location. He does not identify his sensation by some brain event's location.

Like 'snowstorm', 'mutiny', 'flicker', and other words for public events, our names for private events reflect an interest in a range of an event's properties, not only its causal relations.

² A person might look to Wittgenstein and reason: 'An experience's interior properties are incommunicable. We classify experiences solely by outward signs — public causes and effects. Wittgenstein has shown this.' I criticize this argument elsewhere [7].

We name and so specify mental states primarily by their introspectable a-causal properties. A mature natural language is bound to contain names for many of these mental states.

V. Three Introspectable Properties by Which We Specify Experiences

We specify some sensations in part by *temporal* properties. 'Ache' and 'twinge' resemble 'drought' and 'flash' by specifying in part by duration. To be a drought a dry spell must span more than an evening, and a sensation must last longer than a second to be an ache. (A one second back pain may be the beginning of a backache, but it cannot be a 'backache' *simpliciter*.) Like a flash or throb, a twinge or pang is brief. No twinge lasts all day. We individuate aches from twinges, at least in part, by duration.

We also specify various cognitive and perceptual events partly by their duration. No one 'ruminates' for a split second. A person does not 'glance' at his watch for ninety minutes, though he may repeatedly do so throughout this period.

Second, we specify some sensations in part by their felt *location*. With 'backache', 'earache', 'migraine', 'angina', and various others words we do this.

Some people think a sensation's felt location is illusory in some way. ('Think of phantom limb pain', a person may say. 'This sensation cannot be where we experience it as being.') Suppose the location is illusory in some way. We do, nonetheless, experience sensations as having bodily location. That we experience sensations as having a location underlies our general practice of locating sensations and our particular practice of labelling some sensations 'backaches' and others 'headaches'.

Third, we specify many mental states in part by their *unpleasant* or *pleasant* qualitative character or 'quale'.

Unpleasantness is a defining feature of 'pain' and so of 'backache' and other pain-words.

VI. An Experience's Duration and Felt Location are A-causal Properties

An event's location is not a cause or an effect of that event. Nor is its duration.

We cannot explain what it is for an object to have a particular location and duration by referring to its causes and effects. The *reverse* may be necessary. To explain what it is for something to cause an explosion, or be its effect, we may need to specify its spatio-temporal relation to the explosion.

Though the two often coincide, pain's felt location is not itself the location of some injury or other bodily disorder causing the pain.³ Sometimes the pain's location is different from the bodily disorder's. (Sometimes an earache is caused by a diseased tooth.) Sometimes there is no injury. When a person experiences phantom limb pain, there is no bodily part, and hence no physical process, at the pain's site.

³ Gilbert Ryle asks, 'By what criteria do we come to locate or mis-locate sensations . . . ?' Ryle answers this question in part by noting, 'The pain is in the finger in which I see the needle . . .'

The correlation between injury site and pain location is contingent. We benefit from it. The correlation may be a consequence of selection. We put less weight on a leg that hurts. Doing so helps the leg heal. Suppose a child's right foot hurt when he injured his left. He might be slower to favour or otherwise attend to the injured foot.

Pain's felt location is distinguishable from a site a pain disposes a stricken agent to shield or rub. I do not shield or rub my head when I have a headache.

A pain's felt location is not the sensation's property of causing the sufferer to *think* of a particular location. It is not that, aside from the thoughts that accompany each, a sensory experience of ankle pain is, the same as one of shoulder pain. An octopus might have a headache and not think of his head. He may not know what a head is or know he has one. When a person thinks he has a headache, his thought (normally) is founded in his pain's being so located.

A pain may have specific consequences in virtue of having a particular location. (Foot pain may prompt a person to limp. Shoulder pain rarely, if ever, provokes this behaviour.) In calling pain's location an 'a-causal property' I do not deny this. The location is an 'a-causal property' in the sense that it is not itself a cause or an effect of that sensation but a different kind of property the sensation has.

VII. Pain's Unpleasantness is an A-causal Property Intrinsic to Pain

That unpleasant or pleasant quale which is a defining property of pain and pleasure is specifiable, not ineffable. A detailed, exact, non-circular explanation of what it is for a quale to be unpleasant or pleasant follows.

Pain's unpleasantness is a partial cause of our antipathy to pain. The unpleasantness is not, however, pain's property of causing antipathy.

Pain is 'unpleasant' because, qualitatively, pain is intrinsically *bad*, bad in itself and independently of its effects [8].

We partially explain what it is for pain to be intrinsically 'bad' by identifying pain's connection with *reasons* for action. There is sound intrinsic reason to dislike and avoid pain. As Putnam writes, 'X's hurting my hand is *ipso facto* a reason for drawing my hand away from X' [16, p.419].

Pain's intrinsic badness is an a-causal property of pain. 'Pain is intrinsically bad' entails 'There is intrinsic reason for sentient beings to avoid pain' not 'Pain causes sentient beings to avoid it'. Pain is not only unwanted but undesirable (bad).

That pain is intrinsically bad, that there is reason to dislike and avoid pain, does not hinge on our disliking pain. The bad underlies our dislike and is present when the dislike is not. Lewis implicitly acknowledges this. He imagines a man who is not interested in preventing pain or getting rid of it. Lewis assumes this person is not only different from you and me but *mad* and hence irrational. Were there nothing bad in a particular pain and no reason to want to be rid of it, it would not be irrational or 'mad' to be unconcerned about it.

All pain is unpleasant, and all unpleasantness is qualitatively bad [see 8]. Not all unpleasant sensations stimulate aversion. Lobotomized people and masochists are

not averse to all unpleasant sensations.

That we dislike pain is a consequence of pain's being intrinsically bad and giving reason for dislike. Our intelligence and associated sensitivity to reason are partly responsible for our aversion to pain [6].

Unpleasant sensations have other properties in addition to the abstract, formal property of being intrinsically bad. Each unpleasant sensation has a specific introspectable character that is intrinsically bad. Unpleasant sensations differ from each other in their particular character. The unpleasant sensation of smelling sulphur differs qualitatively from the unpleasant sensation a hornet's sting causes. Each of these qualitatively different sensations is 'unpleasant' because it has a quale that is, in its own way, intrinsically bad and so reason giving.⁴

Since unpleasantness is a particular a-causal property of an experience, and unpleasantness is a defining property of 'pain', a particular a-causal property is a defining property of 'pain'. Not just any quale will suffice for pain.⁵

We can embrace mad pain and Martian pain in a single definition of 'pain' without engaging in awkward, *ad hoc* manoeuvres. A definition that focuses on pain's introspectable, intrinsic a-causal properties smoothly embraces mad pain, Martian pain, and, indeed, every other kind of pain.

Pain and pleasure are qualitative opposites through an a-causal property intrinsic to each experience's introspectable interior.

'Pain' and 'pleasure' are not substance neutral. We can know *a priori* that what is not intrinsically bad, qualitatively, is not pain.

A one second, wholly pleasant sensation in my toe would not be a headache whatever its causes and effects or other accompaniments.

Any functionalist principle that affirms that we identify an experience exclusively by its accompaniments is mistaken for many of the reasons that the causal principle is mistaken.

⁴ Pat and Paul Churchland say no single quale is common to pain occasioned by an electric shock, a blow to a knee, a severe burn, and a dull ache [3, p.126]. Perhaps, the particular quale of each of these sensations differs in some way from that of each of the others. Nevertheless, in each case the quale has that abstract property — intrinsic badness — in virtue of which a pain is 'unpleasant'.

⁵ Sidney Shoemaker suspects pain's unpleasant quality is essential to pain. Candidly, he admits this points to what is in fact a fatal objection to the kind of reasoning he uses to defend functionalism from qualia objections.

Shoemaker believes that for visual experience a person can consistently embrace functionalism and admit that qualia inversion is possible. When seeing blue a person must experience 'some qualitative character or other' but no quale in particular. The sensation you have 'seeing blue', I might have 'seeing red', Shoemaker notes [24, p.194]. Shoemaker thinks this admission harmonizes nicely with the view that we identify the experience of 'seeing blue' by its causes and effects.

On the topic of pain and qualia inversion Shoemaker [24, p.162] is tempted to reason (correctly) as follows: Pain does not admit of qualia inversion. Pain, necessarily, is intrinsically unpleasant. A sensation cannot be pain and so be intrinsically unpleasant for one person and not be so for another person. A sensation qualitatively identical to one person's pain must be unpleasant for him and for anyone else. Though no particular quale is necessary for seeing blue, a particular quale — an unpleasant one — is necessary for feeling pain. Not just any quale is sufficient for pain and unpleasantness. This, Shoemaker realizes, does not harmonize nicely with the view that a sensation is 'pain' solely by reason of its causes and effects.

VIII. Two of These A-causal Properties Are Not Orthodox Neural Properties

Paul Churchland observes that the mental states and properties that introspection reveals appear 'radically different from any neurophysiological states and properties' [2, p.29]. Daniel Dennett says that what introspection presents *seems* strikingly different from the 'alien, inscrutable world of billions of brain processes going on in our skulls' [4, p.196]. David Armstrong notes that 'it is perfectly plain' that we are not even aware of an experience 'as material' in introspection [1, p.158].

Suppose an experience were a neural event and that event had nothing but orthodox material properties. In introspection we would be directly aware of a neural event and nothing else. This neural event would have no properties other than orthodox material properties to present. A person would be directly aware of orthodox material properties of brain events most of his life. (We are aware of sensations and other elements in experience most moments in our waking lives.) Why would the world of brain processes seem 'alien' and 'inscrutable'? How could it? Why would a mental state's properties appear 'radically different' from a neural state's? What would prevent us from being aware of an experience as neural in introspection? Why would we not even be aware of it 'as material'?

What we are aware of introspectively does not present orthodox material properties to us. Nor does it present only negative properties, the absence of material properties. It presents positive properties that are different from orthodox material properties.

The rule is: If it is introspectable, it is not an orthodox material property of some neural event.

These two a-causal properties of a backache, its felt location and hedonic quality, are not orthodox material properties of brain events. I will explain.

The word 'backache' is a testament to the fact that backaches have *some* immediate tie to the back. Most people are as confident that their backaches are in their backs as they are that their knuckles are parts of their hands.

A person's brain is in his head. Orthodox physical properties of events in his brain have their immediate, direct spatial tie to his head. The property a backache has in its immediate, direct tie to the back has no perfect duplicate in the class of orthodox material properties of brain events.⁶ A person who identifies a backache's felt location with something which is not a perfect duplicate identifies this property with what it is not.

It is not only that the way we *speak* of a backache's location differs from the way we *speak* of a neural event's. A backache has a kind of location that differs from a neural event's orthodox physical location.⁷ This is *why* our talk of the former's

⁶ This leads Bede Rundle to think materialists should not identify pain with a brain process but with 'the state of the body at or around where the pain is felt' [19, p.184].

⁷ When people use location talk to locate sensations, they use location words with the same sense as when they locate public objects. This is so even if we agree back pain is not *in the back* in the way a bone is. Suppose I say my cat is asleep 'on the stair'. I might be referring to stairs in a house, a dollhouse, a picture, or a dream. The phrase has the same meaning in all of these contexts. 'In the back' is like 'on the stair' in this respect. Its meaning is the same in both the public and private contexts I mention here.

location differs from our talk of the latter's.

Neuroscientists might discover a complex neural *correlate* of pain's being felt in the back. They might discover one or more orthodox material properties of brain events that *cause* a person to feel back pain. To discover this cause or other correlate is not in itself to discover that a sensation's property of being in the back is that orthodox material property.

People will think a philosopher with a rudimentary knowledge of the brain is not in a position to know whether a mental state's properties are identical to orthodox material properties of brain events. They will think only neuroscientists can determine if there is such an identity. Many people take a scientific principle like 'Water is H₂O' as their model of an identity relation. They assume scientists discovered that water is 'identical to' H₂O. People think non-scientists could neither have known nor ruled out this identity *a priori*. Materialists will think that as scientists discovered that water is identical to H₂O, neuroscientists might discover that a backache's felt location is identical to an orthodox material property of a brain event.

We can know *a priori* that neuroscientists will not discover some fact about a mental state's nature that parallels the discovery people represent with the formula 'Water is H₂O'. Water is a material substance. Like other material substances, water is composed of atoms. 'Water is H₂O' is a formula that represents water's *molecular composition*. (What materialists regard as the 'is' of identity is often the 'is' of composition.) Water is composed of water molecules. Each water molecule contains two hydrogen atoms for each oxygen atom. In discovering the truth of 'Water is H₂O' scientists discovered that water has this composition.

A sensation's property of being 'in my back', like a shadow's property of being 'to the right of my leg', is a relation between two objects. This *relation* is not a substance. It has no atomic composition and hence no molecular composition that empirical research might one day disclose.

An ache's introspectable a-causal property of being *unpleasant* — its having a quale that is intrinsically bad and which thus gives reason for avoidance — is not an orthodox physical property of a neural event.

Neuroscientists do not routinely identify neural events as intrinsically bad. An event's intrinsic badness, its giving reason for avoidance, is not the sort of property that might interact with neural properties according to physical principles neuroscientists recognize.

Any attempt to identify a quality's intrinsic badness with some nonrelational, intrinsic material property of a brain event fails by equating a higher order, abstract, supervening property with what is at most a candidate for a lower order, concrete, 'subvening' property. A person who identifies a quality's intrinsic badness with some abstract, formal orthodox material property of a brain event proposes that the normative is identical with something non-normative, which it is not.

Pain's badness is related to voluntary, purposeful action in a way no orthodox neural property is related to neural and bodily events. We may credit voluntary, deliberate pain avoidance behaviour to pain's intrinsic badness. In being bad, pain grounds desires, decisions, and actions aimed at reducing or eliminating pain.

IX. A Role for 'Analyticity'

'A twinge has a short duration' has the logical status of 'All bachelors are unmarried' and 'All vixens are foxes', what Hilary Putnam terms 'trivial examples of analyticity' [16, p.40]. 'Pain is unpleasant', 'A headache is experienced as being in the head', 'If Rachel knows p, p is true', and many other sentences about sensations and cognitive states are also *analytic*.

'A twinge has a short duration' has the epistemic status of 'Every kilometre is a thousand metres'. A person can know the sentence is true simply by understanding it.⁵

My belief that a twinge has a short duration is not merely a well-confirmed empirical generalization. Nature's laws might change such that unsupported objects no longer fall. No change in its laws would lead twinges to acquire the duration of aches and *vice versa*.

In conceiving of a person, horse, or Martian with a twinge I conceive of him experiencing a brief sensation.

'Some twinges last hours' is internally inconsistent. It is like 'Some vixens are bulls'.

In the way being a fox is essential to a vixen, and having four sides is essential to a square, having a short duration is essential to a twinge.

X. Psychological Words Have Many Kinds of Defining Features

Analytic functionalists are monists. Choosing from an experience's diverse properties, they pick one kind (causal relations) to mention in definitions. I am a pluralist. I refer to several kinds of properties when defining psychological words.

An experience has more than one kind of a-causal property. Its qualitative character is only *one* of the a-causal properties functionalists miss.

Some a-causal defining properties of mental states are not introspectable. A partial condition of *knowing* p is that p is *true*. Suppose p is '9 x 9 = 81'. P's being true is an a-causal property of my knowing p. (P's being true is not a cause or an effect of my knowing p.) Its being true is not an introspectable property of my knowing p.

A mental state's qualitative character is only one of its introspectable properties. As there are properties other than colour in what we see (depth, motion, shape), so there are properties other than qualitative complexion in an experience's introspectable interior (duration, location, and others). An experience's qualia — unpleasant or otherwise — present only one dimension of an experience's intro-

⁵ Quine sees a criterion of 'analyticity' in this. Quine proposes that a sentence is analytic when 'to learn even to understand it is to learn that it is true' [17, p.79]. 'This obviously works for "No bachelor is married" and the like . . .', Quine observes [18, p.270]. Various sentences about headaches and other mental states I place in Quine's 'and the like' class.

Quine thinks few sentences are analytic in this way. 'Elementary logic and the bachelor example are clear enough cases, but there is no going on from there', he writes (18, p.271). Various observations in the present paper prove there is 'going on from there'.

spectable interior.

A sensation's duration and felt location are distinguishable from its quale.

A two second itch need not differ qualitatively from a one second itch.

A neck itch need not differ qualitatively from a toe itch. A headache's location is not an intrinsic property of some quale of a sensation. As William James observes, 'No single *quale* of sensation can, by itself, amount to a consciousness of position . . . ' [10, p.154].

I have mentioned four kinds of a-causal properties mental states have — a particular location, duration, hedonic character, and truth value. Each of these a-causal properties is a defining property of some mental states. Mental states have other kinds of a-causal defining properties in addition to those I mention here.

A person cannot define 'car' by specifying only the sounds cars make. Similarly, a person set on specifying only an experience's causes and effects cannot — however refined his description — correctly define 'headache'.

XI. Empirical Functionalism Rests on a False Assumption

Not every grouping of objects, events, states, or properties forms a class whose members have a Kripkean essence we may unearth only through an empirical discovery. Many words do not name Kripkean natural kinds [23, p.19].

An event is a 'headache' by satisfying conditions (a minimum duration, etc.) we can express with various analytic sentences.

Kripke and Putnam insist that statements assigning a natural kind its identifying properties are not analytic. Both developed their theory of natural kinds as an *alternative* to an analyticity-centred view of an object's identity.

Kripke says a word is not a 'rigid designator', and so not a natural kind term, if its meaning and reference is 'given by a description' [11, p.57]. The meaning of 'headache' and other psychological terms can be specified with a description. The conjunction of properties definitive of a 'headache' fix the reference of the word 'headache' each time we use the word.

Instead of being Kripkean natural kind words, psychological words are what Stephen Schwartz [22, p.572] calls *nominal kind terms*. 'Cube', 'lawyer', 'headache', and other nominal kind terms have a sense that determines what the word names.

A class of objects we specify with a nominal kind term N does not have, *qua* N, a 'species essence' we can know of only *a posteriori*.

Like a cube, a backache does not have a species essence that is knowable only *a posteriori*. A backache has no species essence that lies in some set of its causes and effects.

The sense of an exact nominal kind word marks out properties strictly necessary for, or essential to, the class of objects the word names. A backache *qua* 'backache' has a nominal essence. A particular felt location, duration, and hedonic character are essential to headaches and backaches. Of 'headache', 'cube' and other exact nominal kind words we may say, with Wittgenstein [27, #371], 'Essence is

XII. Functionalists Err in What They Include, Not Only in What They Omit

People may say: 'Functionalists need not be silent on the a-causal properties you cite. We can cite a headache's introspectable a-causal properties in a definition specifying a headache's causes and effects. With these amendments we recognize the a-causal properties you mention and preserve analytic functionalism.'

This amendment does not *preserve* analytic functionalism. A-causal properties remain defining features of 'headache' and other psychological words.

Furthermore, few of the causes and effects people include in causal definitions are true defining properties.

At most, *some* psychological words are *partially* definable by reference to a cause or an effect.*

Were analytic functionalists right and a mental state's bodily causes and effects defining features for all psychological words, parallelism and epiphenomenalism would be internally inconsistent. These doctrines are counterintuitive, but they are not internally inconsistent. 'Mind and body do not interact' differs from a statement like 'Frostbite is not caused by exposure to cold' which is internally inconsistent.

Consider this standard functionalist proposal. Jerry Fodor [5, p.118] depicts a headache as:

(That which) causes a disposition for taking aspirin in people who believe aspirin relieves a headache, causes a desire to rid oneself of the pain one is feeling, often causes someone who speaks English to say such things as 'I have a headache', and is brought on by overwork, eye-strain and tension.

Imagine finding *this* in your dictionary as a definition of 'headache'! *None* of the causes and effects Fodor mentions is a defining property of the word 'headache'. *None* is essential to a headache.

What does not attend every headache is not necessary for a headache. None of the causes and effects Fodor mentions attends every headache.

Most of the causes and effects Fodor cites attend only a fraction of headaches. 5% of headaches may be caused by eye-strain. 5% may lead a sufferer to announce his headache.¹⁰

Perhaps, *most* headaches dispose sufferers to desire to be rid of them. Not *all* do.

Nor is it essential to an event's being a headache that it be a *type* of event that is sometimes (or typically) caused by eye-strain or the other conditions Fodor mentions. Eye-strain might *never* cause headaches. Eye-strain may be correlated with

* Perhaps, we assign a sensation a cause when calling it an 'afterimage'. We say the sensation was caused by the sight of some object. A visual sensation caused by a blow to the head rather than by seeing an object is not an 'afterimage'.

¹⁰ When offering functionalist definitions, *many* people mention causes and effects that attend a fraction of cases. (To do so is the norm, not the exception.) A moan, wince, and a feeling of panic are among the effects Paul Churchland suggests may be essential to pain [2, pp.3, 36, 59]. How often do *you* moan, or panic, or wince when you feel pain? Prolonged, intense pain leads people to moan. Most pain is not both intense and prolonged. I suspect fewer than 2% of pains prompt moaning. What attends one pain in fifty is *essential* to pain?

some other condition that causes headaches. (If we discovered this, we would not conclude that people do not have 'headaches'.) Parallel discoveries are possible for 'overwork' and other causes and effects Fodor mentions.

Fodor proposes a conditional link between headaches, belief, and behaviour. His proposal does not come close to giving a necessary condition for a headache. Some headaches do *not* cause a 'disposition for taking aspirin in people who believe aspirin relieves a headache'. A person may not have aspirin. He may prefer ibuprofen. He may spurn orthodox medicine. He may be paralyzed. He may not even mind his ache. Short of modifying Fodor's sentence to read, in effect, 'A headache causes a person . . . to take aspirin *except when it does not*', I see no way of qualifying his proposal to produce an exceptionless, necessary condition for a headache.

Nor is there some determinate *disjunction* of causes and effects an event must satisfy to be a headache.

Suppose I have a sensation with a headache's location, duration, and unpleasant quality. I realize the sensation has none of the causes and effects on Fodor's list. I need not conclude this sensation is not a 'headache'. Other disjunctions of causes and effects people might propose would fail as Fodor's does.

A person can sensibly say he has no idea what caused his 'headache'. Were 'headache' like 'sunburn', and a person picked a condition out by its cause when referring to it with this term, he could not sensibly say this.¹¹

'Headache' is like 'cube' in this respect: with both words we specify an object exclusively by its a-causal properties.

I suppose every headache has *some* cause. Perhaps, all headaches have effects. (People who reject determinism qualify their endorsement of this. They think a headache does not 'cause' the voluntary behaviour it inspires.) Our confidence that headaches have causes and effects does not rest on some specific insight into headaches or the meaning of the word 'headache'. That confidence is founded in the general belief that whatever happens has causes and effects.

Ironically, 'ache' *excludes* rather than includes specific causes. In referring to my pain as an 'ache' I imply that it is *not* sustained by direct corporeal contact with some external agent. Suppose my eight year old causes and sustains pain in my back by pressing her knee against my back. Assume the pain begins with her knee pressure and ends when she withdraws her knee. The pain has a backache's duration, felt location, and unpleasant quality but is not a backache.

Some headaches have public causes and behavioural consequences. These private events are 'headaches' by reason of their introspectable a-causal properties, not their public causes and behavioural consequences.

A person seeking correct definitions of psychological words need not dwell long on functionalism.

Though a causal analysis of experience avoids some objections that refute behaviourism (and invites objections behaviourism does not), *both* views are incorrect.

¹¹ We might substitute 'thought', 'depression', 'dream', and most other psychological words for 'headache'. In each case a person says something coherent when he denies having any idea of what caused the experience to which he is referring.

Lewis says his causal hypothesis 'can be tested, in principle, in whatever way any hypothesis about the conventional meanings of our words can be tested' [13, p.213]. The hypothesis is now tested and refuted.

XIII. Psychological Words are Precise

Fodor does not see specific causes and effects every headache has. Sidney Shoemaker does not see exceptionless causes and effects for a belief [25, p.92]. When people define psychological words functionally, they treat psychological words as cluster terms. Putnam [16, p.328] and Lewis [13, p.212] explicitly claim psychological words are cluster terms.

As their behaviourist forerunners did, functionalists present a *disjunction* of properties as their definition. Fodor is not suggesting every headache is caused by overwork, eye-strain, *and* tension. He is proposing, at most, that every headache has one or more of the causes and effects he mentions. Functionalists and behaviourists view psychological words in the way Wittgenstein saw 'game'.

Like behaviourists, functionalists regard psychological words as *loose*. Had people explicitly affirmed that psychological words are *precise*, few people would have been drawn to functionalism and behaviourism. There are few candidates for specific causes, effects, or behavioural reactions that every pleasure, thought, or other kind of experience has.

Were psychological words loose in the way functionalists assume they are, no specific properties would be individually necessary, and jointly sufficient, for a headache or other mental state. An event would meet the conditions necessary and sufficient for being a 'headache' by having some flexible combination of causes and effects in some assortment of causes and effects.¹² Psychological words would be inexact. When a person says he has a 'headache', he would say, at most, that he has a sensation with some flexible set of properties. He would not say exactly which of these properties the sensation has.

Some people think *all* philosophically interesting words are inexact. 'No interesting descriptive term has any interesting necessary and sufficient conditions', writes Richard Rorty [20, p.307].

There are various conditions necessary for a headache. *Every* headache has *all* of a headache's defining properties — its felt location, minimum duration, unpleasant quality, and independence from concurrent sustaining external agents. Refinement of these necessary conditions will yield conditions that are jointly suffi-

¹² People have not explicitly recognized, and appreciated the implications of, the functionalist's commitment to disjunctive analyses.

People routinely assign experiences 'a causal role' and 'essential' causes and effects. People forget that a disjunctive, family resemblance definition is an *alternative* to one that assigns exceptionless, essential properties. Wittgenstein *contrasts* his family resemblance analysis with essentialism. People who define psychological words disjunctively should deny, not affirm, that an experience has 'a causal role'. (This expression suggests uniformity at the point where the functionalist portrays irregularity.) They should deny, not affirm, that headaches and other kinds of experiences have essential properties.

The idea that pain or thought has 'a causal role' is a *fiction*. Neither the functionalist nor anyone else has identified specific causes and effects every pain or thought has.

cient for a headache.¹³

Like 'sphere' and 'deduction', 'headache' and other psychological words are *exact* nominal kind terms. We assign an object specific properties when referring to it with an exact nominal kind term. We are correct in referring to an object with that word only if the object has every property we assign it with that word.

XIV. A Final Observation About Analyticity

Appeals to the analyticity of certain statements about experiences support pivotal claims in this paper.

To support the claim that analytic claims are philosophically impotent Putnam cites 'Bachelors are unmarried' and 'A chair is a movable seat for one with a back' [16, pp.36-7]. Putnam regards these analytic statements as philosophically sterile. Putnam's examples seem sterile in part because of their *subject matter*. People have little philosophical interest in bachelors and chairs. On these topics philosophers examine *neither* analytic nor non-analytic statements.

Many questions about the mind grip philosophers. What determines a mental state's identify? What properties are strictly necessary, and together sufficient, for a kind of experience? Are empirical functionalists right in thinking psychological words are Kripkean natural kind terms? On these topics analytic claims may assist inquiries.

XV. Conclusion: A Behaviourist Insight

Functionalism is a descendant of the behaviouristic approaches to the mind-body problem Ludwig Wittgenstein and Gilbert Ryle advanced.

David Lewis says behaviourists were right to think statements linking experience to behaviour 'somehow' contain an element of analyticity. His causal principle embraces this insight, he says [12].

Some statements linking pain and pleasure to behaviour *are* analyticity. However, functionalists and behaviourists mis-identify the analytic statements. Statements connecting 'pain' to dispositions to withdraw, moan, wince, or behave in other particular ways are not analytic.

Two *other* kinds of statements connecting pain to behaviour are analytic.

First, 'There is reason to avoid pain' is analytic. There are many ways to avoid pain. To reduce and so avoid pain a person might favour an injured leg, rub a sore jaw, take aspirin, or do many other things. There are logical connections between 'pain' and these actions *qua* 'actions there is reason to perform'.

Second, some words pick out behaviour by the emotional evaluation it manifests. There is more to 'moaning' than emitting a particular sound. A moan expresses anguish over an event one regards as *bad* in some way. (A father might moan when

he sees the car his teenage son just parked begin to roll down the street towards the river.) Pain is intrinsically bad. 'Intense pain gives some reason to feel the emotion a moan expresses', or some similar sentence, is analytic. 'Groan', 'wince', and 'grimace' are similar to '*moan*' in their ties to emotion, evaluation, and pain.

Limping, taking aspirin, moaning, grimacing, and the other forms of behaviour I mention here are distinct from one another.

When people construct functionalist and behaviourist definitions, they produce a definition like Fodor's. They cite a multitude of irregular, distinct reactions. Different people cite different behavioural reactions. When people define 'pain' this way, they miss that property that unites different pains and makes them instances of a single kind of experience.

We recognize the unity among pain sensations when we look beneath pain's diverse behavioural consequences to the introspectable interior of the sensation that underlies, grounds, and provokes these reactions. Intrinsic to *every* pain is the pain's a-causal property of being intrinsically bad. Pains form *a single kind of sensation* by reason of this a-causal property intrinsic to each sensation's introspectable interior.¹⁴

Davidson College

Received March 1992

Revised June 1993

¹³ A headache is *within*, not on the surface of, the head. With this addition, conditions jointly suf-

¹⁴ Exchanges with David Armstrong, Errol Bedford, John Biro, Angela Curran, John Heil, Michael Levin, David Lewis, Bill Lycan, Al Mele, Jack Smart, Cliff Stancil, Bill Throop, and many other people nourished this study. This Journal's (anonymous) referees contributed

REFERENCES

1. D.M. Armstrong and N. Malcolm, *Consciousness and Causality* (Oxford: Blackwell, 1984).
2. P.M. Churchland, *Matter and Consciousness* (Cambridge: MIT Press, 1988).
3. P.M. Churchland and P.S. Churchland, 'Functionalism, Qualia, and Intentionality' in J. Biro and R.W. Shahan (eds) *Mind, Brain, and Function* (Norman, OK: University of Oklahoma Press, 1982) pp.121-145.
4. D. Dennett, 'Review of J. Searle's *The Rediscovery of the Mind*', *The Journal of Philosophy* 90 (1993) pp.193-205.
5. J. Fodor, 'The Mind-Body Problem', *Scientific American* 244 (1981) pp.114-123.
6. I. Goldstein, 'Why People Prefer Pleasure to Pain', *Philosophy* 55 (1980) pp.349-362.
7. I. Goldstein, 'Mental Events and Communication', *American Philosophical Quarterly* 22 (1985) pp.331-338.
8. I. Goldstein, 'Pleasure and Pain: Unconditional, Intrinsic Values', *Philosophy and Phenomenological Research* 50 (1989) pp.255-276.
9. H. Jacoby, 'Empirical Functionalism and Conceivability Arguments', *Philosophical Psychology* 2 (1989) pp.271-282.
10. W. James, *Principles of Psychology* (New York: Henry Holt, 1890).
11. S. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980).
12. D. Lewis, 'An Argument for the Identity Theory', *The Journal of Philosophy* 63 (1966) pp.17-25.
13. D. Lewis, 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50 (1972) pp.249-258 reprinted in Ned Block (ed.), *Readings in Philosophy of Psychology* (Cambridge, MA: Harvard University Press, 1980) vol. 1, pp.207-215.
14. D. Lewis, 'What Experience Teaches' in W. Lycan (ed.), *Mind and Cognition* (Cambridge: Blackwell, 1990) pp.499-519.
15. D.H. Mellor, 'Nothing Like Experience', *Proceedings of the Aristotelian Society* 93 (1993) pp.1-16.
16. H. Putnam, *Mind, Language and Reality*, vol. 2 (Cambridge: Cambridge University Press, 1975).
17. W.V. Quine, *The Roots of Reference* (La Salle, IL: Open Court, 1973)
18. W.V. Quine, 'Two Dogmas in Retrospect', *Canadian Journal of Philosophy* 21 (1991) pp.265-274.
19. B. Rundle, *Perception, Sensation, and Verification* (Oxford: Clarendon Press, 1972).
20. G. Ryle, *The Concept of Mind* (Harmondsworth: Penguin, 1949).
21. R. Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979).
22. S. Schwartz, 'Putnam on Artifacts', *The Philosophical Review* 87 (1978) pp.566-574.
23. S. Schwartz, 'Natural Kinds and Nominal Kinds', *Mind* 89 (1980) pp.182-195.
24. S. Shoemaker, *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984).
25. S. Shoemaker and R. Swinburne, *Personal Identity* (Oxford: Blackwell, 1984).
26. M. Tye, *The Metaphysics of Mind* (Cambridge: Cambridge University Press, 1989).
27. L. Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Blackwell, 1953).