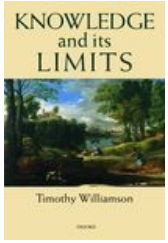


Knowledge and its Limits

Timothy Williamson

https://doi.org/10.1093/019925656X.001.0001

Published: 2002 **Online ISBN:** 9780191598678 **Print ISBN:** 9780199256563

CHAPTER

3 Primeness

Timothy Williamson

https://doi.org/10.1093/019925656X.003.0004 Pages 65–92

Published: October 2002

Abstract

A composite condition is one that consists of the conjunction of a purely internalist condition and a purely external condition; a condition that is not composite is prime. A general method of argument is provided for showing that many mental states, including knowing, are prime. A connection is made with the account of knowledge that Plato gives in the *Meno*. The primeness of mental states is shown to contribute to their value in the explanation of action by facilitating a kind of generality, which is made precise by use of the probabilistic notion of a correlation coefficient. The account is extended to the explanation of non-mental events.

Keywords: action, composite, condition, correlation coefficient, explanation, externalism, generality, internalism, Meno, Plato, prime

Subject: Epistemology, Metaphysics

Collection: Oxford Scholarship Online

3.1 Prime and Composite Conditions

The previous chapter argued, in a preliminary way, that internalism is false and that, on the externalist alternative, knowing is a genuine mental state. This chapter deepens the critique of internalism. It argues on structural grounds that the envisaged separation of internal and external factors is impossible. Many ordinary mental states are not equivalent to conjunctions of something purely internal with something purely external. This inequivalence is essential to the causal-explanatory work which their attribution can do. The internal does not play the distinctive role in the explanation of action that internalism predicts. On an externalist understanding of mental states, knowing is a central exemplar.

Here is a sketch of an internalist line of thought:

> The causing of my present action is here and now. Only narrow conditions supervene on the here and now; so narrow conditions must play a privileged role in the causal explanation of action. If a causal explanation of action cites a broad mental condition, an underlying narrow condition must do the real work. We can isolate that narrow condition by subtracting from the broad mental condition the environmental accretions that make it broad. We can recover the original broad mental condition from the narrow condition by adding back those accretions.[1]

The internalist conceives the original broad mental condition as the conjunction of the narrow condition and a condition as purely external as the former is purely internal—for instance, the condition that one believes truly that it is raining as the conjunction of the narrow condition that one believes that it is raining and the environmental condition that it is raining.

p. 66 Let us state the matter more formally. Recall that a case α is *internally like* a case β if and only if the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β. A condition C is *narrow* if and only if for all cases α and β, if α is internally like β, then C obtains in α if and only if C obtains in β. C is *broad* if and only if it is not narrow. We can define external likeness on the model of internal likeness: a case α is *externally like* a case β if and only if the total physical state of the external environment in α is exactly the same as the total physical state of the external environment in β. A condition C is *environmental* if and only if for all cases α and β, if α is externally like β, then C obtains in α if and only if C obtains in β. In other terminology, environmental conditions supervene on or are determined by the physical state of the external environment. A condition C is *composite* if and only if it is the conjunction of some narrow condition D with some environmental condition E. As a special case, a narrow mental condition is trivially composite, for it is the conjunction of itself with the environmental condition that holds in all cases whatsoever. C is *prime* if and only if it is not composite. The line of thought that began with the here-and-nowness of causation led to the conclusion that mental conditions are composite.

That internalist line of thought is inconclusive, not least because it uses ill-defined notions of adding and subtracting conditions. The next section will show its conclusion to be false; many of the mental conditions which we attribute to each other in ordinary language are prime. It begins with a preliminary exploration of primeness, compositeness, and some related notions.

3.2 Arguments for Primeness

A broad mental condition entails various narrow conditions. Indeed, there is a strongest narrow condition which it entails, that is, a narrow condition which it entails and which entails every such narrow condition. To see this, let *virtual-C* be the condition which obtains in a case α if and only if C obtains in some case internally like α. Virtual-C is narrow because internal likeness is transitive and symmetric. C entails virtual-C because internal likeness is reflexive. Moreover, virtual-C entails every narrow condition which C entails; for if C entails a narrow condition D, and virtual-C obtains in a case α, then C obtains in some case β internally like α, so D obtains in β (since C entails D), so D obtains in α (since D is narrow); hence virtual-C entails D.
p. 67 Thus virtual-C is the ↳ strongest narrow condition which C entails. When C is a broad mentalistic condition ascribed in natural language, internalists regard virtual-C as the purely mental reality underlying C. In particular, when C is the condition that one knows *p*, they are tempted to identify virtual-C with the condition that one believes *p*, or the condition that one rationally believes *p*. Section 2.3 argued that those identifications are incorrect.

We can define the dual notion of the weakest narrow condition that entails C by substituting 'every' for 'some' in the definition of 'virtual-C'; it obtains in a case α if and only if C obtains in every case internally like α. However, the resulting condition will usually be impossible when C is broad. For what case α does the

condition that one believes that tigers growl obtain in every case internally like α? Virtual-C, the strongest narrow condition which C entails, is the condition of interest to the internalist.

Given a condition C, there is also a condition—call it *outward-C*—which stands to the external as virtual-C stands to the internal. Outward-C obtains in a case α if and only if C obtains in some case externally like α. Just as virtual-C is the strongest narrow condition which C entails, so outward-C is the strongest environmental condition which C entails.

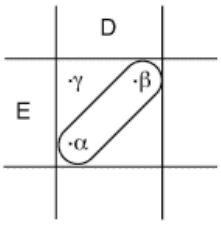
We can now identify narrow and environmental conditions of which a given condition, if composite, is the conjunction: they are virtual-C and outward-C respectively. If C is any conjunction of narrow and environmental conditions at all, then it is the conjunction of virtual-C and environmental-C. We can prove that as follows. Let C be the conjunction D ∧ E of a narrow condition D and an environmental condition E. Since C entails D, virtual-C entails D; similarly, outward-C entails E. Thus the conjunction of virtual-C and outward-C entails D ∧ E, that is, C. Conversely, C entails the conjunction of virtual-C and outward-C, whether or not C is composite. Consequently, C, if composite, is the conjunction of virtual-C and outward-C. To argue that C is prime is in effect to argue that C can fail to obtain when both virtual-C and outward-C obtain.

How can we show that a condition C is prime? Suppose that C obtains in two cases α and β. Consider a case γ internally like α but externally like β; we may assume that such a case is possible because otherwise that interdependence of the internal and the external would itself undermine the idea that they can be separated (see section 3.3 for more on this assumption). Now suppose that C is the conjunction of a narrow condition D with an environmental condition E. Then C obtains in γ. For since C entails D, D obtains in α; since D is
p. 68 narrow, D also obtains in γ, which is internally like α. Similarly, since C entails E, E ↳ obtains in β; since E is environmental, E also obtains in γ, which is externally like β. Since C obtains whenever both D and E obtain, and they both obtain in γ, C obtains in γ, as required. Thus we can show that C is prime simply by exhibiting three cases α, β, and γ, where γ is internally like α and externally like β, and C obtains in α and β but not in γ. We shall see below how to do that for most ordinary mental conditions.

Conversely, the condition C is composite if no such triple of cases exists, for then C obtains in γ whenever both virtual-C and outward-C obtain in γ, so the conjunction of virtual-C and outward-C entails C, and the converse entailment is automatic. Thus it is necessary as well as sufficient for C to be prime that it obtains in two cases but not in a case internally like one and externally like the other.

A picture may help (Fig. 1). The horizontal axis represents total internal physical states; the vertical axis represents physical states of the external environment. The point representing γ is in the same position on the horizontal axis as the point representing α (because γ is internally like α) and in the same position on the vertical axis as the point representing β (because γ is externally like β).

Figure 1



The area between the two vertical lines represents a narrow condition; the area between the two horizontal lines represents an environmental condition. The rectangle formed by their intersection represents a composite condition. The area enclosed by the curve represents a prime condition.

A structural analogy may also clarify what is going on. Suppose that a property P is the conjunction of a colour property Co and a shape property Sh, and that both a black sphere and a white cube have P. Then a black cube also has P: it has Co because it is the same colour as the black sphere, which has Co, and it has Sh because it is the same shape as the white cube, which has Sh. By contraposition, if a black sphere and a white cube have a property which a black cube lacks, then that property is not the conjunction of a colour property and a shape property.

p. 69 Given a mental state S, how can we find three cases α, β, and γ of the ↳ required kind to show that the condition that one is in S is prime? We can construct them to a common pattern. We imagine circumstances in which S can be realized in just two ways, which need not be mutually exclusive. One is in S if and only if one is in S in either way 1 or way 2. Each way involves a channel with an internal and an external part; one is in S in way *i* if and only if both the internal and the external parts of way *i* are open (at this level of simplification, we may treat the condition that one is in S in way *i* as composite). In case α, both the internal and the external parts of way 1 are open but neither the internal nor the external part of way 2 is open; thus one is in S in way 1 although not in way 2; therefore one is in S. Case β reverses the two ways in status. In β, neither the internal nor the external part of way 1 is open but both the internal and the external parts of way 2 are open; thus one is in S in way 2 although not in way 1; therefore one is in S. In case γ, the internal part of way 1 but not the internal part of way 2 is open, because γ is internally like α, and the external part of way 2 but not the external part of way 1 is open, because γ is externally like β; thus one is in S in neither way 1 (because its external part is not open) nor in way 2 (because its internal part is not open); therefore one is not in S. The relations between α, β, and γ ensure that the condition that one is in S is prime. This structure can be represented diagrammatically (Fig. 2).

Figure 2

Case	Way	Internal	External	Joint	Result
α	1	✔	✔	✔	✔
	2	✘	✘	✘	
β	1	✘	✘	✘	✔
	2	✔	✔	✔	
γ	1	✔	✘	✘	✘
	2	✘	✔	✘	

Thus, as a first example, we can argue that the condition that one sees water is prime. Let α be a case in which one sees water normally with one's right eye. One's left eye receives light rays that by chance are like those it would receive from water, but they are all emitted by a waterless device just in front of that eye;
p. 70 however, a head injury prevents further processing of input from one's left eye. Let β be a case ↳ which differs from α by reversing the roles of the two eyes. In β, one sees water normally with one's left eye. One's right eye receives light rays that by chance are like those it would receive from water, but they are all emitted by a waterless device just in front of that eye; however, a head injury prevents further processing of input from one's right eye. Now consider a case γ internally like α and externally like β. In γ, a head injury prevents further processing of input from one's left eye, because it does so in α, and γ is internally like α. Equally, in γ, one's right eye does not receive light rays from water, because it does not do so in β, and γ is externally like β. Thus, in γ, neither eye both receives light rays from water and has its input to the brain subject to further processing. Consequently, in γ, one does not see water. Yet, in α and β, one does see water. By the earlier argument, the condition that one sees water is prime. Obviously, for almost any *x*, the example can be modified to show that the condition that one sees *x* is prime; it is not the conjunction of a narrow condition and an environmental condition.

That the example exploits the binocularity of vision is inessential. We could make the same point by supposing that in α there is water on the right and gin (which looks just like water) on the left, and a brain lesion causes one visually to register only what is on the right. In β there is gin on the right and water on the left, and a brain lesion causes one visually to register only what is on the left; in the case γ internally like α and externally like β, there is gin on the right and water on the left (as in β), and the brain lesion causes one visually to register only what is on the right (as in α). Thus, given appropriate background conditions one sees water in α and β but not in γ. This example is consistent with monocular vision.

For an aural analogue, suppose that, in α, Mary emits sound waves only of frequency *f* while John emits sound waves only of frequency *g*, and a brain lesion causes one aurally to register sound waves only of frequency *f*. In β, John emits sound waves only of frequency *f* while Mary emits sound waves only of frequency *g*, and a brain lesion causes one aurally to register sound waves only of frequency *g*. In the case γ internally like α and externally like β, John emits sound waves only of frequency *f* while Mary emits sound waves only of frequency *g* (as in β), and a brain lesion causes one aurally to register sound waves only of frequency *f* (as in α). Thus, given appropriate background conditions one hears Mary in α and β but not in γ. Examples of this type can be constructed for the other senses.

We can demonstrate the primeness of the condition that one believes that this screen flickers by substituting this screen for water in the first example. For we can assume that in α and β one's belief that
p. 71 this screen ↳ flickers concerns this screen only in virtue of one's visual link to this screen; in γ, since one fails to see this screen, one lacks the belief. The point generalizes to other object-dependent contents and other propositional attitudes.

Consider the condition that one believes that tigers growl. Let α be a case in which tigers inhabit the mountains while schmigers (which appear just like tigers) inhabit the jungle; one remembers one's encounters with tigers in the mountains but totally forgets one's encounters with schmigers in the jungle. One believes that tigers growl; since one has no recollection of schmigers, one does not believe that schmigers growl. Let β be a case in which tigers inhabit the jungle while schmigers inhabit the mountains; one remembers one's encounters with tigers in the jungle but totally forgets one's encounters with schmigers in the mountains. One believes that tigers growl; since one has no recollection of schmigers, one does not believe that schmigers growl. Now consider a case γ internally like α and externally like β. In γ, tigers inhabit the jungle while schmigers inhabit the mountains; one remembers one's encounters with schmigers in the mountains but totally forgets one's encounters with tigers in the jungle. One believes that schmigers growl; since one has no recollection of tigers, one does not believe that tigers growl. Thus the condition that one believes that tigers growl is prime.

We can make the example more vivid by supposing that one believes propositions by storing sentences in a language of thought which express them in a belief box, although it should be clear that nothing in the underlying structure of the example really requires a language of thought or a belief box. In each case, encounters with animals of a suitable appearance in the mountains cause one in a standard way to use a word T1 in one's language of thought (if at all) as a natural kind term for those animals. Encounters with animals of a suitable appearance in the jungle cause one in a standard way to use a word T2 in one's language of thought (if at all) as a natural kind term for those animals. One's language of thought also has a word G that means *growl*. In α, T1 means *tigers* and has the appropriate causal connections with tigers; one believes that tigers growl because one stores the sentence T1G in one's belief box. T2 does not mean *tigers*, because it lacks the appropriate causal connections with tigers; one does not store the sentence T2G in one's belief box. β differs from α by reversing the roles of T1 and T2. In β, T2 means *tigers* and has the appropriate causal connections with tigers; one believes that tigers growl because one stores T2G in one's belief box. T1 does not mean *tigers*, because it lacks the appropriate causal connections with tigers; one does not store T1G
p. 72 in one's belief box. In γ, one does not store T2G in one's belief box, because one does ↳ not do so in α and γ is internally like α. Equally, in γ, T1G does not express the proposition that tigers growl, because T1 does not

mean *tigers*, since T1 lacks the appropriate causal connections with tigers in β and γ is externally like β. Thus, in γ, neither T1G nor T2G both expresses the proposition that tigers growl and is stored in the belief box. We can legitimately assume that in none of the three cases does any sentence in the language of thought other than T1G and T2G express the proposition that tigers growl. Consequently, in γ, one stores no sentence which expresses the proposition that tigers growl in one's belief box; one therefore fails to believe that tigers growl. Yet, in α and β, one does believe that tigers growl. Again, the point generalizes to other externally individuated contents and other propositional attitudes.

We can argue that epistemic conditions are also prime. The previous example might even suffice, for in some cases α and β of the specified kind one *knows* that tigers growl; in α, one fails to know that tigers growl because one fails to believe that tigers growl. However, it is more illuminating to pick an example in which the belief is held constant and what varies is its epistemic status. Let α be a case in which one knows by testimony that the election was rigged; Smith tells one that the election was rigged, he is trustworthy, and one trusts him; Brown also tells one that the election was rigged, but he is not trustworthy, and one does not trust him. Let β be a case which differs from α by reversing the roles of Smith and Brown; in β, one knows by testimony that the election was rigged; Brown tells one that the election was rigged, he is trustworthy, and one trusts him; Smith also tells one that the election was rigged, but he is not trustworthy, and one does not trust him. Now consider a case γ internally like α and externally like β. In γ, one does not trust Brown, because one does not trust him in α, and γ is internally like α. Equally, in γ, Smith is not trustworthy, because he is not trustworthy in β, and γ is externally like β. Thus, in γ, neither Smith nor Brown is both trustworthy and trusted. We can legitimately assume that in none of the three cases does one have any other way of knowing that the election was rigged. Consequently, in γ, one does not know that the election was rigged. Yet, in α and β, one does know that the election was rigged. Thus the condition that one knows that the election was rigged is prime. Since the example does not turn on the specific content of the knowledge, it can be modified to show for almost any proposition *p* that the condition that one knows *p* is prime.

Endless examples can be constructed to the foregoing pattern. Henceforth, mental conditions will therefore be assumed to be characteristically prime. They are not conjunctions of narrow conditions and environmental conditions.

p. 73

3.3 Free Recombination

When we construct triples of cases α, β, and γ to establish the primeness of a mental condition, the possibility of a case like γ depends on the principle that given cases α and β, there is a case internally like α and externally like β. Call that principle *free recombination*. It allows us to treat the internal and the external in a sense as independent variables.

Free recombination is not wholly unproblematic. If the internal and the external are nomically connected, then γ might violate nomic constraints even though α and β do not. For example, if determinism holds and the external includes the past (as it must for the treatment of issues about reference), then the external nomically determines the internal. Although a nomically impossible case might not be metaphysically impossible, such a case would not show very much, for if mental conditions coincided with conjunctions of internal and external conditions across all nomically possible cases that would be a significant vindication of the internalist picture of the mind. Moreover, the internal and the external are constitutively interdependent in other physical ways too. They are supposed to cover mutually exclusive and jointly exhaustive spatial regions; thus the region occupied by one determines the region occupied by the other. When the spatiotemporal interface between the internal and the external is contoured differently in α and β, mismatches threaten the construction of γ. Variations in the physical state of one include variations in the shape of the region it occupies, and therefore constrain variations in the physical state of the other.[2]

Nevertheless, free recombination may still hold to a first approximation, just as colour and shape can be treated to a first approximation as independent properties of an object, even if its colour ultimately depends on its microscopic geometry. We might handle the interdependence of the internal and the external just noted by minor modifications of the framework, for example, by restricting what aspects of the past count as environmental. Furthermore, counterexamples to free recombination are really a problem for the attempt to separate the internal and external under attack in this chapter. It is therefore fair to assume recombination in criticizing that attempt.

Consider, for example, Jerry Fodor's claim, 'identity of causal powers has to be assessed *across* contexts, not *within* contexts' (1987: 35), which he uses in defence of an individualistic conception of the mental (he has
p. 74 subsequently changed his position in his 1994). The proposal is ↳ roughly that individuals have the same causal powers if and only if for every context *c*, they can do the same things in *c*. Of course, an individual's causal powers depend on its internal state, which must therefore be held fixed while the context varies. Thus a more precise formulation of Fodor's proposal is that an individual in an internal state I in some context has the same causal powers as an individual in an internal state J in a possibly different context if and only if for every context *c*, an individual in I in *c* can do the same things as an individual in J in *c*. Suppose that in case α one is in internal state I_α and context c_α; in case β one is in internal state I_β in context c_β. Are one's causal powers the same in α and β? By Fodor's criterion, a necessary (but insufficient) condition for sameness is that one can do the same things in I_α in c_β as one can do in I_β in c_β. This comparison breaks down, independently of what one can do in any case, unless one can be in I_α in c_β. But to be in I_α in c_β is to be in a case γ internally like α and externally like β, for internal likeness is sameness in one's internal state and external likeness is sameness in one's context. Thus the relevant comparisons can be made only if free recombination holds. Where recombination fails, Fodor's test does not grant sameness of causal powers, no matter what one can do in any case. Suppose, for instance, that any case internally like α and externally like β is discounted because it would violate the nomic constraints relative to which causal powers are being assessed; then one would not count as having the same causal powers in α and β. The same result follows if a case internally like α and externally like β is impossible because the spatio-temporal interface between the internal and the external is contoured differently in α and β. Yet these grounds for withholding the verdict of sameness of causal powers do not even mention what one can do; intuitively, they are inadequate. Fodor's test serves its purpose only if free recombination holds. Since his test is implicit in the internalist picture of mental causation, that picture requires free recombination. Thus an argument against the internalist picture can legitimately assume free recombination, for without it the picture fails anyway.

The dialectical position is similar for more general doubts about the distinction between the internal and the external. We should not assume without argument that subatomic physics will embody locality principles of a kind that would guarantee a clearcut distinction between those features which contribute to internal physical states and those which do not. The arguments for primeness are not restricted to mental conditions; they apply to other sorts of physical condition too (see also section 3.7). We may therefore find unexpected difficulty in defining the initial set of unproblematically internal physical conditions. That
p. 75 would ↳ destabilize the very distinction between the internal and the external. Such destabilization is bad news for the internalist conception of the mental under attack in this chapter, for that conception depends on separating the contributions of narrow and environmental conditions, which makes only as much sense as the distinction between the internal and the external itself does. If the distinction is unclear, the externalist can still insist on the negative point that no clarification of it counts ordinary mental conditions as composite. The present conception of the mental does not require a clearcut distinction between the internal and the external; we merely grant such a conception to its internalist rivals for the sake of argument, and then show that, even so, mental conditions cannot be decomposed into narrow and environmental conditions as they envisage.

3.4 The Explanatory Value of Prime Conditions

Do concepts of prime conditions serve any theoretical purpose? In the examples that demonstrate primeness, what is the point of classifying case γ separately from cases α and β? This section argues that we need concepts of prime conditions for the same reasons for which we need concepts of broad conditions generally.

Consider seeing. What is the point of classifying a case in which one sees water separately from a case in which one is in exactly the same internal physical state but sees only a mirage? The difference may not matter for one's action at the next instant, if the action is itself individuated by its internal physical nature. If it were individuated broadly, we should still be wondering about the point of broad individuation. But our interest is not confined to action at the next instant—if there is one; if time is dense, there is not. One is thirsty; how likely is one to be drinking soon? Likely enough, if one sees water. Much less likely, if what one sees is a mirage. Even if drinking is individuated narrowly, its explanation in terms of the earlier state of the system involves the presence of water in the environment, not just the earlier internal physical state of the agent. Concepts of broad mental conditions give us a better understanding of connections between present states and actions in the non-immediate future, because the connections involve interaction with the environment (see also Burge 1986b and Peacocke 1993 on broad explanations of action).

The need to think about connections between earlier mental states and later actions is largely a need to
p. 76 think about connections between ↳ mental states and actions separated by seconds, days, or years. The causal explanation of action is frequently concerned with the structure of the agent's deliberation. But deliberation frequently occurs some time before the moment of action. In deliberating, one assesses alternative courses of action in the light of one's beliefs and desires, decides which is best, and forms the intention to pursue it; one puts the intention into effect only when the time for action comes. How and whether one puts the intention into effect depend on one's interaction with the environment in the intervening period. At the moment of action, one may not even remember one's deliberations in any detail. To confine the explanation of action to the instant before action is to omit much of what makes action rational. Historical explanation is certainly not confined to the instant before action; 'Why did Napoleon invade Russia?' is not a question about his state an instant before the invasion began, after months of planning.[3] Moreover, most actions take time; one cannot instantaneously eat an apple, write a letter, or go for a walk. Extended actions involve complex interaction with the environment.

Could we analyse each action into basic physical actions, and then explain each basic action in terms of the agent's internal state at the preceding instant? Conjoining those proximal explanations of the basic actions would not yield a good explanation of the original non-basic action. Suppose, for example, that we wish to explain why someone went for a walk. Perhaps we can analyse the walk into a sequence of steps. But for each step, the proximal explanation of his taking it will not mention what explains why he went for a walk: that he desired exercise.

For the reasons for which we need concepts of broad conditions, we need concepts of prime conditions. The relevance of seeing water now to drinking soon is not exhausted by the agent's internal state and the presence of water. Before one can drink the water, one must get oneself to it. Typically, one will steer one's way by keeping the water in sight and making a complex series of adjustments to one's position in a feedback loop. The present coincidence of one's internal physical state with the state in which one would be if one saw the water from that perspective is not enough; the coincidence must continue until one reaches
p. 77 the water. The kind of causal relation in which one stands to something ↳ when one sees it often enables one to keep it in sight. By contrast, if the matching of internal state and external environment is mere coincidence, then there is no reason why it should continue. We can find just this contrast in the cases which demonstrated the primeness of the condition that one sees water. Other things being equal, one can keep

the water in sight with one's right eye in case α and with one's left eye in case β; in case γ, one has no means of maintaining the match between internal state and external environment. The match might continue, but that would be good luck. Even if it does continue, that is not seeing; whether one sees now does not depend like that on what happens in the future. Thus the need to understand the connection between present states and action in the non-immediate future gives us reason to classify case γ separately from cases α and β. To classify in that way is to use a concept of a certain prime condition.

We should not expect such considerations to yield a *definition* of seeing. As already noted, attempts to provide non-circular necessary and sufficient conditions for ordinary concepts have a miserable record of failure. The concept of seeing is the resultant of very many forces. The forces considered here make the concept of seeing a concept of a prime condition; they need not determine that condition uniquely.

The argument generalizes to other prime conditions. Sometimes, the content of a propositional attitude depends on a perceptual link to an object, as when one believes that this screen flickers; the foregoing considerations apply immediately. At other times, our thought of an individual or kind depends on a causal link independent of present perception. Nevertheless, such a link is typically *renewable*: it enables us to have further causal interactions with the individual or kind, or at least further causal dependencies on it if it no longer exists, for example, by finding out more about it and acting on that information—none of which implies that reference could be defined in terms of renewable causal links. Of the cases that demonstrated the primeness of the condition that one believes that tigers growl, such renewability is likely to be available in α and β but not in γ. In each case, one can test one's attribution of growling by going to the places where (some of) the creatures to which one attributes it were encountered; that takes one to the habitat of tigers in α and β and to the habitat of schmigers in γ. In encountering them there, one renews the causal link. If tigers really do growl while schmigers do not, one's belief is likely to survive in α and β but not in γ (we may suppose that in each case the tigers or schmigers were not growling when encountered but looked disposed to growl). Thus the need to understand the connection between present states and action in the non-
p. 78 immediate future gives us reason to use concepts of ↳ conditions that are prime because the content constitutively depends on a causal link with an individual or kind.

We can extend the argument to knowledge in more detail, taking a hint from Plato. In the *Meno* (97A–98A), Socrates raises a question about the value of knowledge. Knowing that this road goes to Larissa is useful to you if you want to go to Larissa. But merely believing truly that this road goes to Larissa seems to be equally useful to you, for you will get there just the same. Why should we value knowledge more than mere true belief? Socrates responds by a comparison with the statues of Daedalus, which run away unless they are tethered. True beliefs are liable to be lost, unless they are so anchored that they constitute knowledge.[4]

What does Plato mean? Surely he recognized that mere true beliefs can be held with dogmatic confidence, and knowledge lost through forgetting. But belief can also be sensitive to evidence. One can lose a mere true belief by discovering the falsity of further beliefs on which it had been essentially based; quite often, the truth will out. One cannot lose knowledge that way, because a true belief essentially based on false beliefs does not constitute knowledge. For example, I might derive the true belief that this road goes to Larissa from the two false (but perhaps justified) beliefs that Larissa is due north and that this road goes due north; when dawn breaks in an unexpected quarter and I realize that this road goes south, without having been given any reason to doubt that Larissa is due north, I abandon the belief that this road goes to Larissa. Since that true belief was essentially based on false beliefs, it did not constitute knowledge. The case is an obvious variation on Gettier's counterexamples to the analysis of knowledge as justified true belief.[5]

In other cases, a true belief not essentially based on false beliefs still fails to constitute knowledge, because misleading evidence against that true belief is rife in one's environment, although one happens to be unaware of it oneself. For example, I might correctly classify a dog by sight as friendly; in ordinary circumstances I might thereby come to know that it is friendly. However, this one behaves in ways which, if

observed, would justify the false suspicion that it is hostile. So far it happens to have behaved like that only
p. 79 when my back was turned, and I have not yet formed any such suspicion. But my true belief that it is ↳ friendly does not constitute knowledge, and could be lost at any moment. To know that it is friendly, I must not be surrounded by such misleading counterevidence, and my true belief must not be too vulnerable to this kind of overturning. The case is a variation on Harman's examples of the undermining of knowledge by evidence one does not possess (Harman 1973: 143–4 and 1980: 164–5; see also Goldman 1976: 772–3).

Present knowledge is less vulnerable than mere present true belief to *rational* undermining by future evidence, which is not to say that it is completely invulnerable to such undermining. If your cognitive faculties are in good order, the probability of your believing *p* tomorrow is greater conditional on your knowing *p* today than on your merely believing *p* truly today (that is, believing *p* truly without knowing *p*).[6] Consequently, the probability of your believing *p* tomorrow is greater conditional on your knowing *p* today than on your believing *p* truly today.[7] Of course, profoundly dogmatic beliefs which are impervious to future evidence and do not constitute knowledge may be even more likely to persist than beliefs that are rationally sensitive to future evidence and do constitute knowledge, but then the subject's cognitive faculties are not in good order. Since the difference between your present knowledge and your present true beliefs matters for predicting your future beliefs, it matters for predicting your future actions, because they will depend on your future beliefs.

The evidence which may undermine mere present belief needs time to emerge. As argued in section 2.3, the difference between knowledge and belief can make a present psychological difference; for instance, knowledge excludes various kinds of irrationality that belief does not. If C is the condition that one knows *p*, virtual-C can fail in several ways to be the condition that one believes *p*. However, the present argument concerns only delayed impact, not action at the 'next' instant. We do not value knowledge more than true belief for instant gratification.

We should not expect to define knowledge in terms of persistent true belief, still less in terms of subsequent action. What the argument does suggest is that when a condition stated in non-circular terms (belief, truth,
p. 80 justification, causation, . . .) fails to be necessary and sufficient for ↳ knowledge, that divergence will yield a divergence in implications for future action; the task of stating non-circularly a condition equivalent to knowledge with respect to implications for future action is no easier than the task of stating non-circularly a condition necessary and sufficient for knowledge. On the view defended in chapter 1, both tasks are impossible. Consequently, the mental state of knowing makes a distinctive contribution to the causal explanation of action.

These considerations apply to the cases that demonstrated the primeness of epistemic conditions. In α and β, one knows *p* by testimony, and one's belief in *p* is correspondingly stable; one has also been told *p* by someone untrustworthy whom one distrusts, but that does not make one's belief less stable, for it does not rest on that testimony. In γ, one believes *p* because one trusts the untrustworthy informant who tells one *p*; one distrusts the trustworthy informant who tells one *p*. One's belief is less stable, for in the long run one may recognize the untrustworthiness of the informant on whom one relies. That one may also recognize the trustworthiness of one's other informant is only partial compensation. The description of γ involves a specific threat to one's belief. The descriptions of α and β involve no such specific threat; although one might come to distrust the trustworthy informant, no reason has emerged why one should. On mildly cheerful assumptions about normal background conditions, there is at least some tendency for the truth on such matters to out, so one's belief in *p* is more stable in α and β than in γ. The need to understand the connection between present states and action in the non-immediate future gives us reason to use concepts of prime epistemic conditions.

3.5 The Value of Generality

Some may nevertheless claim that prime conditions are theoretically redundant. For let α be any case in which a prime condition C obtains, and D and E respectively the strongest narrow condition and the strongest environmental condition which obtain in α. Then it is plausible that the conjunction D ∧ E entails C. For D completely specifies the internal physical state of the subject, and E completely specifies the physical state of the rest of the world, so, unless some physical relations fail to qualify as either internal or external (for example, for holistic reasons), D ∧ E completely determines the physical state of the world in α; and if the total state of the world supervenes on its total physical state, then D ∧ E entails C. Although the
p. 81 assumptions just made are not ↳ uncontroversial, let us allow them for the sake of argument. Thus, in any particular case, all the consequences we want of a prime condition follow from a compound condition obtaining in that case. Why should our best theory bother with the prime condition?

Let F be the condition that one subsequently performs a certain action. Suppose that F obtains in α; we want to know why. If the prime condition C makes F highly probable, given background conditions, we are tempted to cite C in explaining why F obtains. But D ∧ E presumably makes F certain, so why not cite D ∧ E rather than C? Doesn't it give the real causal explanation?

The answer to the would-be rhetorical questions is this. Our best theory is intended to capture significant generalizations. The action would have been performed in many cases other than α, in which D ∧ E does not obtain; D ∧ E is sufficient but nothing like necessary for F. A theory which relies on conditions like D ∧ E may leave uncaptured a significant generalization relating F to C. What has not been shown is that significant generalizations about prime conditions can be replaced by significant generalizations about compound conditions.

Good explanations have an appropriate generality. If one cites a sufficient condition for the condition to be explained, or one near enough so for the purpose in hand, the purported explanation can nevertheless fail because the condition to be explained would still have obtained in the same way even if the cited condition had not obtained. For example, one can explain why someone died by saying that he was run over by a bus; the explanation becomes worse, not better, if one specifies that the bus was red, for its colour had nothing to do with his death. If all metals have a certain property, one will be unhappy with attempts to explain why gold has it which cite properties of gold not shared by other metals. Again, if a condition obtains necessarily, then to explain why it obtains by deriving it from conditions which obtain only contingently is to miss its modal generality (as conventionalist explanations characteristically do).[8]

Many features of the maximally specific condition D ∧ E will be quite irrelevant to the obtaining of F. They
p. 82 will concern physical events that ↳ form no part of the causal chain between the agent's initial mental state and the final performance of the action. The agent would have performed the action anyway, even if those features had been different. Their inclusion is a defect in the explanation. A highly specific account may constitute a good explanation of *how* something happened without constituting a good explanation of *why* it happened.[9]

Reductionist strategies of explanation risk providing bad explanations by citing highly specific conditions and thereby missing the generality of the conditions to be explained. Successful reductions involve no loss of generality. Something common to all genuine instances of the given phenomenon is identified in lower-level terms. The present argument does not undermine the explanatory value of those reductions. But that value is not shared by explanations which use no such generalization about the phenomenon, and merely provide—or rather gesture towards—a maximally specific description in lower-level terms of the particular case at hand.[10]

Defences of narrow content often treat the total state of the external environment ('circumstances', 'context') as one component of the favoured explanations of action, the other comprising attitudes to narrow contents. On the face of it, this is to give up generality across different states of the environment. Yet such accounts do not allow the internal component of the explanation to be similarly unarticulated; the condition that one has a certain attitude to a certain 'narrow content' is supposed to be a general one, obtaining in a range of different cases. Once generality is acknowledged as an explanatory virtue, the question arises whether it can best be achieved by explanations that factorize in the envisaged way.

We need concepts of prime conditions to achieve appropriate generality in explaining action. Consider again the cases that demonstrated primeness: α and β were mutually symmetric; the best explanation of the agent's subsequent actions might well generalize across α and β, citing a condition that obtains in both. But if it also obtained in γ, the explanation would be weakened, for then the cited condition would not rule out a range of cases in which the agent's subsequent actions in α and β are much less likely (see section 3.4). Thus the cited condition should obtain in α and β but not in γ, and therefore be prime.

p. 83 Of course, generality is only one of many explanatory virtues. Some ↳ purported explanations achieve spurious generality by using disjunctive concepts. For example, if someone was crying because she was bereaved, it does not improve the explanation to say that she was crying because she was bereaved or chopping onions. But ordinary mental concepts of prime conditions (such as the concept of seeing) are not disjunctive (see also section 1.5). To argue that such concepts do not express genuine common properties on grounds of explanatory uselessness would be viciously circular, for such concepts give generality to our explanations. Unless they are already assumed not to express common properties, nothing has been done to undermine their apparent explanatory usefulness.

When we explain why a condition C obtains by citing a prior condition D, the generality of our explanation varies inversely with P[C|~D], the probability of C conditional on ~D; in that sense, we lack generality to the degree to which D fails to be necessary for C.[11] The converse explanatory virtue is sufficiency, which varies with P[C|D], the probability of C conditional on D. We can combine these into a single explanatory virtue, the degree to which C is *correlated* with D. The higher the correlation, the better the answer to the question 'Does D obtain?' as a guide to the answer to the question 'Will C obtain?'; correlation is also a predictive virtue.

Correlation is itself only one of many explanatory virtues, but it is the one of present interest. A more rigorous framework for discussing it will be expounded, and then applied to the use of prime conditions in the causal explanation of action.

3.6 Explanation and Correlation Coefficients

In probability theory, the standard measure of correlation is the *correlation coefficient*, which takes values between +1 (perfect positive correlation) and −1 (perfect negative correlation; Appendix 1 gives technical details). Formally, this coefficient measures the correlation between random variables, which themselves
p. 84 take numerical values. For present purposes, what matter are correlations between conditions. We can ↳ adapt the standard concept in a natural way to speak of the correlation coefficient of two conditions by associating each condition with its indicator random variable, which takes the value 1 when the condition obtains and 0 otherwise, and defining the correlation coefficient of two conditions as the correlation coefficient of their associated indicator random variables. The coefficient α[C,D] of correlation between the conditions C and D can then be calculated in terms of their probabilities and that of their conjunction. The result is a slightly unperspicuous formula:

$$\frac{P[C|D] - P[C]P[D]}{\sqrt{(P[C](1 - P[C])P[D](1 - P[D]))}}$$

The probabilities here are objective properties (chances) of the conditions, for the degree to which two conditions are correlated is an objective matter. But they are not single-case probabilities, for conditions are general, like properties; they can obtain in many actual cases. The relevant probability space comprises both actual and merely possible cases; they will be circumscribed by a set of background conditions, which vary with the explanatory context.

One can easily check that C and D are positively correlated ($\rho[C,D] > 0$) if and only if the probability P[C|D] of C conditional on D exceeds the unconditional probability P[C] of C (Appendix 1, proposition 2): one condition raises the probability of the other. The conditions are perfectly positively correlated ($\rho[C,D] = 1$) if and only if P[C|D] = 1 and P[C|~D] = 0 (Appendix 1, proposition 5): C is certain to obtain if D obtains and certain not to obtain if D does not obtain, so whether C obtains can be predicted with certainty on the basis of whether D obtains.

In explaining action, our concern is with imperfect positive correlations ($0 < \rho[C,D] < 1$). For example, the condition that one will perform a certain action in the future may be imperfectly positively correlated with both the condition that one presently knows some proposition and the condition that one believes that proposition truly (we can treat the relevant desires as background conditions). The question was: with which of the latter two conditions is the former condition better correlated? More generally, which of two conditions D and E, both positively correlated with C, is more highly correlated with C? It turns out that $\rho[C,D] \leq \rho[C,E]$ if and only if

$$(P[C|D] - P[C])(P[C] - P[C|\sim D]) \quad \leq \quad (P[C|E] - P[C])(P[C - P[C|\sim E])$$

p. 85 (Appendix 1, proposition 3). Thus both the degree to which D helps us ↳ to predict C (P[C|D] P[C], the degree to which D raises the probability of C) and the degree to which ~D helps us to predict ~C (P[C] P[C|~D], the degree to which ~D lowers the probability of C) are relevant. In particular, if E raises the probability of C more than D does and ~E lowers the probability of C more than ~D does, then C is better correlated with E than with D (Appendix 1, proposition 4).

A completely schematic example may clarify the picture. We can take the example of knowledge, and develop the account in section 2.4 of its explanatory value. Similar points can be made about other prime conditions. Let C be the condition that one will perform a certain action, D the condition that one believes truly a proposition *p*, and E the condition that one knows *p*. The background conditions may include the agent's desires and other beliefs. Suppose that just three equiprobable possibilities have non-zero probability: one believes *p* truly and knows *p* and will perform the action; one believes *p* truly without knowing *p* and will not perform the action; one neither believes *p* truly nor knows *p* and will not perform the action. Hence $P[C \wedge D \wedge E] = P[\sim C \wedge D \wedge \sim E] = P[\sim C \wedge \sim D \wedge \sim E] = 1/3$. Thus it is certain that one will perform the action if and only if one knows *p*; P[C|E] = 1 and P[C|~E] = 0. The two conditions are perfectly positively correlated; $\rho[C,E] = 1$. It is also certain that one will perform the action only if one believes *p* truly, so P[C|~D] = 0. However, if one believes *p* truly, one may or may not perform the action, depending on whether one knows *p*; P[C|D] = 1/2. These two conditions are imperfectly positively correlated; calculation shows that $\rho[C,D] = 1/2$. Realistic examples have none of this simplicity. But since the correlation coefficient is a continuous function of the relevant probabilities, small enough changes in the latter make small changes in the former. Thus we can introduce some of the messy complexity of real life into the example and still have performing the action better correlated with knowing than with believing truly ($\rho[C,D] < \rho[C,E]$). In

particular, this comparative ranking is consistent with non-zero probabilities for the possibilities that one performs the action without knowing *p*, with or without believing *p* truly (C ∧ D ∧ ~E and C ∧ ~D ∧ ~E) and that one fails to perform it while knowing *p* (~C ∧ D ∧ E). Although in most realistic examples we cannot expect to calculate exact probabilities or correlation coefficients, such comparative rankings can still be plausible. In very crude terms, if the probability of performing the action conditional on knowing *p* exceeds the probability of performing it conditional on believing *p* truly without knowing *p* by much more than the latter exceeds the probability of performing it conditional on failing to believe *p* truly, then performing the
p. 86 action is more highly correlated with knowing *p* than with believing *p* truly. A precise statement of the ↳ principle would take into account the prior probabilities of knowing *p* and believing *p* truly.

Action is often more highly correlated with belief or with true belief than with knowledge. But not always. You see someone coming to your door; he is about to knock loudly. You are tempted not to reply. How would he react? You ask yourself, 'Does he know that I am in?' not, 'Does he believe that I am in?' If before knocking he does know that you are in, then he is unlikely to abandon his belief if you fail to reply; he will probably take offence. If before knocking he believes (truly) without knowing that you are in, then he is much more likely to abandon his belief if you fail to reply; he will probably not take offence. If before knocking he fails even to believe that you are in, then he is even less likely to take offence. Whether he would take offence is better predicted by whether he knows than by whether he believes. His taking offence is more highly correlated with knowing that you are in than with believing (truly) that you are in.

The point is not that knowing exceeds believing in implied degree of confidence; it need not. I know many things without being prepared to bet my house on them. The example works even if the degree of confidence required for belief is stipulated to be the same as the degree of confidence required for knowledge. The visitor who merely believes (truly) when he knocks that you are in may be exceedingly confident that you are in, but abandon the belief when to his astonishment you do not reply, for he is even more confident that if you do not reply you are not in.[12] But someone who knows that you are in has grounds that will not be undermined just by your failure to reply. Clearly, all this is a matter of probability; if your visitor merely believes that you are in, he *may* retain the belief when you do not reply, and take offence; equally, if he knows that you are in, he *may* abandon the belief in a fit of self-doubt when you do not reply. Nevertheless, the probabilities are often as indicated above.

Consider a variation on an example used in section 2.4, this time involving the attribution of beliefs to non-human animals. How long would we expect a fox to be willing to search for a rabbit in the wood before giving up, assuming initially (a) that the fox knows that there is a rabbit in the wood, or (b) that the fox
p. 87 believes truly that there is a rabbit in the wood? In (b) but not (a), the fox's initial true belief may fail to ↳ constitute knowledge because the true belief is essentially based on a false one, for instance, a false belief that there is a rabbit in a certain hole in the wood. When the fox discovers the falsity of that belief, the reason for the search disappears. That will not happen in (a), because a true belief essentially based on a false one does not constitute knowledge. Thus, given plausible background conditions, more persistence is to be expected in (a) than in (b). In many such cases, lengthy persistence is better explained by initial knowledge than by initial true belief.

Sometimes the predictive difference between knowledge and true belief is mediated by the cultural significance of knowledge. A notorious criminal may try to eliminate all those who know that he killed the policeman, because they are potential witnesses against him in court. He will not bother to eliminate those who merely believe truly that he did it, because their confidence that he did it, however great, is no threat to him, given the rules of forensic evidence. If we want to predict whether someone will soon be fleeing for her life, 'Does she know that he shot the policeman?' is a better question than 'Does she believe that he shot the policeman?'. It is better even than the question 'Does she believe that she knows that he shot the policeman?' when flight is contingent on the criminal's behaviour and he believes that the testimony of

anyone who mistakenly believes themselves to know that he did it will not stand up in court. The danger is knowing too much, not believing too much.

We can also use a schematic example to reinforce the conclusion of section 3.5, by showing in detail how a very specific condition strictly sufficient for the condition to be explained can nevertheless fail to be highly correlated with it. Let C be the condition that one will perform a certain action, D the very specific condition obtaining in the case at hand (for example, completely determining the agent's internal physical state and the physical state of the environment), and E the condition that one knows *p*. The background conditions include the agent's desires. Assume that D is sufficient for both C and E, which leaves just five possibilities. Suppose that they have these probabilities:

$$\begin{aligned}
&P[C \wedge D \wedge E] = 1/10\\
&P[C \wedge \sim D \wedge E] = 3/10\\
&P[C \wedge \sim D \wedge \sim E] = 1/10\\
&P[\sim C \wedge \sim D \wedge E] = 1/10\\
&P[\sim C \wedge \sim D \wedge \sim E] = 4/10
\end{aligned}$$

Thus it is certain that one will perform the action if the specific condition obtains (P[C|D] = 1) and not certain that one will perform it if one knows *p* (P[C|E] = 4/5). However, one is much more likely to perform
p. 88 it ↳ if the specific condition does *not* obtain than if one does *not* know *p* (P[C|~D] = 4/9 > 1/5 = P[C|~E]). The latter disparity in favour of E more than compensates for the former disparity in favour of D; calculation shows that ρ[C,D] = 1/3 < 3/5 = ρ[C,E]. Performing the action is better correlated with knowing *p* than with the strictly sufficient specific condition.

3.7 Primeness and the Causal Order

A high correlation does not guarantee a direct causal connection. When the condition that one knows *p* is highly correlated with the condition that one will perform a certain action, the reason might be that both the knowledge and the action are effects of a common cause, without the knowledge causing the action. What would it be for the knowledge not to cause the action? Presumably, the condition that one knows would not be causally relevant in the right sense to the condition that one will perform the action. But then we should not focus on the former in explaining why the latter obtains. Does this seriously threaten the role of prime conditions in the explanation of action?

High correlations are an indispensable though fallible guide to causal structure. Where a high correlation misleads us into falsely postulating a causal connection, more detailed information about further correlations should correct our mistake. The high correlations between prime mental conditions and conditions on subsequent action constitute defeasible evidence for the causal effectiveness of the prime conditions. Higher correlations constituting a genuinely rival explanation would be needed to defeat that evidence.

Given deterministic laws, we might define a present condition D perfectly correlated with the condition C that one will perform the action, by stipulating that D obtains in a case α if and only if the total present state of the system (agent and environment) in α and the deterministic laws entail that one will perform the action. C and D obtain in exactly the same nomically possible cases. Thus, if the laws have probability 1, C and D are perfectly correlated (if P[C] > 0; otherwise ρ[C,D] is ill defined). D is not defined disjunctively. However, the definition of D unifies the cases in which D obtains by what happens later (the performance of the action), not by the present state of the system. In many contexts, such a correlation will not give us the kind of understanding we seek. It certainly does not give us what we need for purposes of prediction and

p. 89 control, but that is not quite the same thing. We seek a correlation ↳ between a condition given by a concept that unifies the cases in which it obtains in terms of the present state of the system and a condition given by a concept that unifies the cases in which it obtains in terms of the future state of the system; we are willing to sacrifice some degree of correlation in order to achieve such unification.

Even if we can replace the conditions conceived by folk psychology by conditions more highly correlated than they are with the condition that one subsequently performs the action, those new conditions will themselves be prime (as D is above), for reasons already indicated. Moreover, such explanations may well constitute refinements rather than refutations of the folk psychological explanations.

Discussions of broadness have tended to concentrate on intentional content. Since intentional content is a mental phenomenon, the causal efficacy of broad conditions, and specifically of prime conditions, can appear to require special pleading on behalf of the mental. It does not. The considerations of this chapter are not confined to the mental. For example, one can demonstrate in the style of section 3.2 the primeness of the condition that a ship is anchored to the seabed; it is not the conjunction of a condition on the internal physical state of the ship and a condition on the physical state of its external environment. Clearly, the condition that a ship is anchored to the seabed can be causally effective with respect to the ship's subsequent motion or rest. Primeness is no bar to causal efficacy. It derives its significance from our interest in causal explanatory connections between states of objects and their subsequent behaviour (in the widest sense) after an interval long enough to permit intervening interaction with their environment. That is the normal case, not the exception, in causal explanation.

3.8 Non-Conjunctive Decompositions

The arguments in section 3.2 for the primeness of various mental conditions were not supposed to show that those conditions cannot be analysed somehow as functions of narrow and environmental conditions. A composite condition is the conjunction of a narrow condition with an environmental condition. How far do the problems identified in this chapter for conjunctive analyses generalize to analyses of other forms?

Conjunction is not the only truth-function. Disjunction is a simple alternative. Call a condition *non-trivial* if and only if it obtains in some cases but not in all. Then we can easily show, given free recombination, that
p. 90 the (inclusive) disjunction of a non-trivial narrow condition with a ↳ non-trivial environmental condition is always prime.[13] Thus an argument for the primeness of a mental condition does not automatically show that it is not such a disjunction.

Of course, given free recombination, we should not expect a mental condition to be the disjunction of a non-trivial narrow condition with a non-trivial environmental condition. For if it were, and the environmental condition obtained in a case β, then for any case α, some case γ would be internally like α and externally like β. Since the environmental condition would obtain in γ, the mental condition would too (because a disjunction is entailed by its disjuncts); and thus the mental condition would be consistent with any non-trivial narrow condition whatsoever (sawdust in the head, . . .), which is implausible. We could also argue that a mental condition C is not the disjunction of a narrow condition D and an environmental condition E by arguing that the contradictory condition ~C is prime, for if C were D ∨ E then ~C would be ~(D ∨ E), which is ~D ∧ ~E; since the contradictory of a narrow condition is itself narrow, and the contradictory of an environmental condition is itself environmental, ~D ∧ ~E is composite. But the possibility that a mental condition is a more complex function of narrow and environmental conditions cannot be dismissed so easily.

Suppose, for example, that a mental condition C involves some kind of matching between one's internal state and the state of the external environment, although it does not fix those states separately. Then a

simple hypothesis would be that C is a possibly infinite disjunction $(D_1 \wedge E_1) \vee (D_2 \wedge E_2) \vee \ldots$, where $D_1, D_2, \ldots$ are narrow conditions, E_1, E2, . . . are environmental conditions, and D_i matches E_i in the appropriate sense. Although each disjunct $D_i \wedge E_i$ is composite, disjunctions of composite conditions are not themselves usually composite. The required matching between internal and external states may occur in cases α and β separately without occurring in a case γ internally like α and externally like β; $D_i \wedge E_i$ and $D_j \wedge E_j$ may each entail matching while $Di \wedge E_j$ does not. Equally, the matching may occur in γ without occurring in α or β, so
p. 91 the disjunction of composite conditions is not even the contradictory ↳ of a composite condition. Thus, more realistically prime conditions can be constructed as quite simple truth-functions of narrow conditions and environmental conditions.

A less unsophisticated proposal is that the mental condition requires some causal relation between one's internal state and the matching state of the external environment. That would only strengthen the argument for primeness. Of course, causal relations to the environment are often conceived as themselves on the external side, in which case they could be subsumed under the environmental conditions; but since they also implicate their internal relata, that conception of them endangers free recombination. The causal relation is better conceived as bridging the internal and the external.

Since a prime condition may be a truth-function or some subtler function of narrow and environmental conditions, the arguments for the primeness of various mental conditions do not show that our concepts of those conditions cannot be analysed into concepts of narrow and environmental conditions. The arguments for unanalysability are different; as in section 1.3, they advert to the long history of failed analyses, the lack of any good reason to expect analysability, and the availability of an alternative understanding of the mental. Nevertheless, the arguments for primeness are needed to fix the role of the mental in the causal explanation of action. For even if a mental condition C were a disjunction $(D_1 \wedge E_1) \vee (D_2 \wedge E_2) \vee \ldots$ of conjunctions of non-trivial narrow conditions D_i with non-trivial matching environmental conditions, it would not follow that C could be replaced in causal explanations by corresponding narrow and environmental conditions; a composite condition can be so replaced. Given free recombination, the strongest narrow and environmental conditions entailed by the disjunction are $D1 \vee D_2 \vee \ldots$ and $E_1 \vee E_2 \vee \ldots$ respectively.[14] But if $(D_1 \wedge E_1) \vee (D_2 \wedge E_2) \vee \ldots$ is prime, then it is not entailed by its composite consequence $(D_1 \vee D_2 \vee \ldots) \wedge (E_1 \vee E_2 \vee \ldots)$. Only the former requires one's internal state to match the state of the external
p. 92 environment. When the causal explanation depends on the primeness of $(D_1 \wedge E_1) \vee (D_2 \wedge E_2) \wedge \ldots$, as ↳ section 3.4 argued that it often will, the extractable narrow condition $D1 \vee D_2 \vee \ldots$ typically plays no explanatory role; it is a sort of epiphenomenon. What would give the narrow condition an explanatory role is compositeness, not analysability; the arguments for primeness therefore tell against such an explanatory role for the narrow condition.

If an explanation specified an environmental condition E_j, we might combine that with the disjunction $(D_1 \wedge E_1) \vee (D_2 \wedge E_2) \vee \ldots$ to derive the corresponding specific narrow condition D_j (if E_j were incompatible with E_i for every *i* distinct from *j*), which then would play a distinctive explanatory role. But specificity is lack of generality; sections 3.5 and 3.6 showed how lack of generality can be an explanatory vice. An explanation at an appropriate level of generality will be neutral between the disjunct $D_j \wedge E_j$ and some alternative disjunct $D_i \wedge E_i$, while still excluding $D_i \wedge E_j$ and $D_j \wedge E_i$; the unspecific narrow condition $\ldots \vee D_i \vee \ldots \vee Dj \ldots$ extractable from that explanation plays no distinctive explanatory role therein. Non-conjunctive decompositions of the mental into narrow and environmental conditions do not save the internalist picture of the mind, for they do not give narrow conditions the explanatory role which it predicts for them.

Notes

1 Terminology is borrowed from Jackson's useful survey (1996), although without attribution of the argument to him.

2 See also Susan Hurley's related critique of what she calls the Duplication Assumption in ch. 8 of her 1998.

3 Consider also Putnam's example (1978: 42): 'Professor X is found stark naked in the girls' dormitory at 12 midnight. Explanation: (?) He was stark naked in the girls' dormitory at midnight –ε, and he could neither leave the dormitory nor put on his clothes by midnight without exceeding the speed of light. But (covering law:) nothing (no professor, anyhow) can travel faster than light.'

4 The passage is briefly discussed by Williams 1978: 38, Shope 1983: 12–13, and Craig 1990a: 7.

5 The passage in the *Meno* is sometimes cited as a source of the view that knowledge is justified true belief; on this interpretation, Plato would be mistaken about the nature of the tether. The aim is not to elucidate his intentions here. See Fine 1992: 218–19 for further discussion and references.

6 Knowing *p* is assumed to entail believing *p*; if not, the claim should be about the conditional probability of knowing or believing *p* tomorrow. Similar adjustments could be made throughout.

7 The argument uses the easily established fact about conditional probabilities that if E entails D and $P[C|D \wedge \sim E] < P[C|E]$ and $P[E|D] < 1$, then $P[C|D] < P[C|E]$. Let C be the condition that you believe *p* tomorrow, D that you believe *p* truly today, and E that you know *p* today.

8 Consider also another of Putnam's examples (1978: 42): 'A peg (1 inch square) goes through a 1 inch square hole and not through a 1 inch round hole. Explanation: (?) The peg consists of such-and-such elementary particles in such-and-such a lattice arrangement. By computing all the trajectories we can get applying forces to the peg (subject to the constraint that the forces must not be so great as to distort the peg or the holes) in the fashion of the famous Laplacian super-mind, we determine that some trajectory takes the peg through the square hole, and no trajectories take it through the round hole. (Covering laws: the laws of physics.)'

9 For example, the explanations envisaged by Jackson and Pettit 1995: 277 suffer from this loss of generality.

10 See also Yablo 1992 and 1997 on proportionality between causes and effects, and Steward 1997: 192–7 on causal relations between facts.

11 The phrase 'degree of necessity' could equally well be associated with other measures, such as P[D|C]. An advantage of a probabilistic analysis is that it forces one to attend to such distinctions. For present purposes, the natural probabilities to consider are those conditional on the prior conditions D and ∼D rather than those conditional on the outcome conditions C and ∼C.

12 In probabilistic terms, if P_{new} is the result of updating P_{old} by conditionalization on the new evidence $\sim r$, then $P_{old}[i]$ can be arbitrarily high and $P_{new}[i]$ arbitrarily low, if $P[i|\sim r]$ is sufficiently low, which will require $P_{old}[r]$ to be high (let *i* be that you are in and *r* that you reply). Note that the relevant interpretation of 'P' here, in contrast with the rest of the chapter, is as degree of belief (credence), not as objective probability.

13 Proof: Suppose that the narrow condition D obtains in case δ but not in case δ*, while the environmental condition E obtains in case ε but not in case ε*. By free recombination, there are cases α, β, and γ, where: α is internally like δ* and externally like ε; β is internally like δ and externally like ε*; γ is internally like δ* and externally like ε*. Thus γ is internally like α and externally like β. Since E is environmental and obtains in ε, which is externally like α, E obtains in α; thus D ∨ E obtains in α. Since D is narrow and obtains in δ, which is internally like β, D obtains in β; thus D ∨ E obtains in β. But since E does not obtain in ε*, which is externally like γ, E does not obtain in γ; since D does not obtain in δ*, which is internally like γ, D does not obtain in γ; thus D ∨ E does not obtain in γ. Therefore, D ∨ E is prime.

14 Proof: $\vee_i(D_i \wedge E_i)$ obviously entails $\vee_i D_i$. Suppose that $\vee_i(D_i \wedge {}^{i}_{E})$ entails a narrow condition D. We must show that $\vee_i D_i$ already entails D; we need only show that an arbitrary D_j entails D. Suppose that D_j obtains in a case α. Since E_j is non-trivial, it obtains in some case β. By free recombination, there is a case γ internally like α and externally like β. Since D_j is narrow and E_j is environmental, they both obtain in γ; thus $\vee_i(D_i \wedge E_i)$ obtains in γ; since it entails D, D obtains in γ; since D is narrow, it also obtains in α. Thus ^{Dj} entails D, as required. Consequently, $\vee_i D_i$ is the strongest narrow condition which $\vee_i(D_i \wedge E_i)$ entails. By a parallel argument, $\vee_i E_i$ is the strongest environmental condition which $\vee_i(D_i \wedge E_i)$ entails.