Mechanisms of mental causation:
An examination of the theories of Anomalous Monism and Direct Realism with regard to their proposals concerning the causal role of human mentality in the natural world.

A thesis submitted in partial fulfilment of the requirements of Doctor of Philosophy,
School of Psychology, Faculty of Science, University of Sydney.

Acknowledgements

Throughout the process of writing the present thesis, I have been so fortunate as to spend many happy hours in discussing realist philosophy with my supervisors Joel Michell and Terence McMullen, as well as with Fiona Hibberd. I should like to take this opportunity to express my gratitude to each of these good friends, and to acknowledge the positive influence that they have had on my learning experience.

I should also like to thank my parents, Faith Charity and Brian Medlow, for their loving support and the encouragement that they have shown me throughout all stages of my education.

Most of all I am grateful to James Palethorpe who has been a fellow student and loving companion throughout the whole.

Sharon Medlow
Abstract

One of the most interesting developments in recent psychological theorising has been a growing appreciation of the need for a viable theory of mental causation. Hitherto, the prospects for reconciling what seems to be the uniquely rational character of human thought and action with the non-rational mechanistic workings of the natural world have appeared to be limited or even illusory, and the pursuit of reconciliation of this sort has therefore formerly been dismissed as being either impossible of completion or inappropriate for contemplation. Much of the scepticism concerning the role of causal processes in human thought and action was dispelled, however, by the philosopher Donald Davidson, who argues that not only is human action capable of being caused by the actor’s thoughts and desires, but that only when such action is so caused, can it be rational.

Davidson’s proposal for the reconciliation of human rationality with causal necessitation is articulated in his theory of Anomalous Monism. According to this theory, there exists what may be termed an ontological-conceptual distinction between events themselves and the characters or properties that are attributed to events by human observers, and it is through recognition of this distinction that one discovers how mental events, that is, events that are amenable to description in the psychological vocabulary, are causally efficacious yet free from the constraints typically associated with the necessity and sufficiency of causal laws. Anomalous Monism, if it were workable, would therefore resolve the paradox according to which human mentality is at once integrated in, and yet unconstrained by, the mechanistic natural world, by demonstrating the compatibility of the facts of causation with the intuitions of folk psychology.

However, close examination of Anomalous Monism reveals it to rely on logically flawed anti-realist principles concerning the characters of events, properties and causation. It follows from this that the theory itself must be rejected, but the task that it was devised to undertake, the formulation of a viable theory of mental causation, need not be similarly discarded. Rather, what remains is the challenge of delineating an alternative theory, one that withstands logical scrutiny whilst addressing what is characteristic of human mental processes, and thereby what is characteristic of mental causation.
The theory of Direct Realism that is derived from the broader philosophical realism of John Anderson provides the materials for meeting this challenge. According to Direct Realism, mental phenomena are relational situations obtaining between certain organisms (including humans) and their environments. As such, mental phenomena are included in the range of phenomena occurring in the natural world and they are therefore subject to all of its ways of working, including its deterministic mechanisms. The particular challenge that a Direct Realist theory of mental causation faces, that of demonstrating that relational situations can be causal, is revealed upon examination of the character of causation to be unproblematic. Furthermore, the seeming incompatibility between human rationality and natural necessitation is resolved when it is acknowledged that, rather than be an inherent feature of thought and action, logical structure is a characteristic of the natural environment that organisms are at times sensitive to, as revealed by its effects on the characters of their thoughts and actions.

Far from being remote or illusory, the prospects for reconciling human mentality with the causal mechanisms of the natural world are discovered in the present thesis to be favourable when a realist approach to the characters of both mental events and causation is adopted.
## CONTENTS

### I: INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.i Human freedom in a determined world</td>
<td>i</td>
</tr>
<tr>
<td>1.ii The purpose of psychology: prediction and control or understanding of meaning?</td>
<td>ii</td>
</tr>
<tr>
<td>1.iii Methodological issues</td>
<td>v</td>
</tr>
<tr>
<td>1.iv Ontological issues</td>
<td>vii</td>
</tr>
</tbody>
</table>

### PART I: DAVIDSON’S THEORY OF ANOMALOUS MONISM

#### Section 1: Establishing the Paradox

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Davidson and rationality</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Davidson and mental anomalism</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1 Strict laws and generalisations</td>
<td>11</td>
</tr>
<tr>
<td>1.2.2 Mental anomalism: why strict laws are absent from psychology</td>
<td>18</td>
</tr>
<tr>
<td>1.2.2.1 Externalism</td>
<td>19</td>
</tr>
<tr>
<td>1.2.2.2 Dispositionality</td>
<td>21</td>
</tr>
<tr>
<td>1.2.2.3 Rationality</td>
<td>23</td>
</tr>
<tr>
<td>1.3 Davidson’s paradox</td>
<td>25</td>
</tr>
</tbody>
</table>

#### Section 2: Resolving the Paradox

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Characterising events</td>
<td>28</td>
</tr>
<tr>
<td>2.1.1 Events as a fundamental ontological category</td>
<td>28</td>
</tr>
<tr>
<td>2.1.2 Events as particulars</td>
<td>31</td>
</tr>
<tr>
<td>2.2 Constitutive description</td>
<td>33</td>
</tr>
<tr>
<td>2.2.1 Theories and their empirical interpretations</td>
<td>34</td>
</tr>
<tr>
<td>2.2.2 Mental attributions</td>
<td>37</td>
</tr>
<tr>
<td>2.2.3 The Unified Theory and radical interpretation</td>
<td>40</td>
</tr>
<tr>
<td>2.3 Causal relations and causal explanations</td>
<td>49</td>
</tr>
<tr>
<td>2.3.1 Causal relations and singular causal statements</td>
<td>50</td>
</tr>
<tr>
<td>2.3.2 Causal explanations</td>
<td>51</td>
</tr>
<tr>
<td>2.4 Physical-mental event token identity</td>
<td>55</td>
</tr>
<tr>
<td>2.5 Supervenience</td>
<td>56</td>
</tr>
<tr>
<td>2.5.1 Supervenience in ethics and aesthetics</td>
<td>57</td>
</tr>
<tr>
<td>2.5.1.1 Subvenient and supervenient properties</td>
<td>57</td>
</tr>
<tr>
<td>2.5.1.2 The standard in supervenience</td>
<td>59</td>
</tr>
<tr>
<td>2.5.2 Mental supervenience</td>
<td>65</td>
</tr>
<tr>
<td>2.5.2.1 Supervenient and subvenient properties in mental supervenience</td>
<td>65</td>
</tr>
<tr>
<td>2.5.2.2 The standard in mental supervenience</td>
<td>68</td>
</tr>
<tr>
<td>2.5.2.3 The conceptual status of supervenient properties</td>
<td>72</td>
</tr>
<tr>
<td>2.6 Resolving the paradox</td>
<td>76</td>
</tr>
</tbody>
</table>
I: Introduction

I.i Human freedom in a determined world

The suggestion that thought and action can be entirely accounted for in terms of the causal mechanisms that operate in the natural world is incompatible with traditional understandings of what it is to think and behave.

The widely held conviction that while, as a matter of fact, one’s past courses of action did unravel in certain ways, they could just as well, with the exertion of one’s will, have turned out differently, was famously illustrated in a lecture delivered by William James in 1884. James asked his audience to imagine that at the close of the meeting he, having the option of walking home via either Divinity Avenue or Oxford Street, proceeded through Divinity Avenue only to have the powers of the universe annihilate ten minutes of time and thereby return him to his starting position, ready to begin his journey home. At this point, he argued, although all circumstances were now as they were ten minutes before, he would be free to vary his past course of action by ignoring Divinity Avenue and walking home via Oxford Street instead.

The appeal of James’ (1884) argument is partly accounted for by the fact that it accords with what, for many of us, goes without question; the view that we are free to think as we please and act in accordance with our whims and fancies. Despite this, there exists good reason to suppose that thought and behaviour are caused, and it is only if this supposition is correct that the practice of psychological experimentation and therapy can be justified.

It is the lot of the psychologist, therefore, to question the freedom of the mind and to either discover a compromise, if such exists, between the liberty of free will and the necessitations of the natural world, or to reject one or the other of these. The past century has witnessed variations on each of these alternatives. Championing a deterministic view have been the behaviourists, with their focus upon the mechanisms by which all natural events, including behaviour, proceed. In contrast of focus to the behaviourists have been the hermeneuticists, who centre their attention upon the rationality of human mental processes as a distinctly human characteristic. What is curious about the hermeneutic and behaviouristic schools is that neither has attempted
to discover a mechanism in which mental events play a causal role. Causation and mental processes, it has been outrightly assumed by both, are mutually exclusive. Accordingly, in order to provide a deterministic account of behaviour, the behaviourists have eschewed investigation of all possible mental causes of behaviour, while the hermeneuticists, in their efforts to understand mental processes, have neglected causal inquiry.

Important attempts to remedy the exclusion of mental phenomena from causal processes, that is, attempts to found a workable theory of mental causation, can be found in the writings of Donald Davidson, who advocates a reconciliation between human agency and causation, and of the Direct Realists, who argue that there is no opposition between mentality and causal necessitation, that is, that mental events are themselves caused and causally efficacious. It is with these latter two positions that the current thesis is directly concerned, but before turning my attention to them, I shall briefly review the behaviouristic and hermeneutic movements with a view to demonstrating why mental events and causal relations have previously been deemed incompatible.

I.ii The purpose of psychology: prediction and control or understanding of meaning?

The behaviourist movement was characterised by its commitment to demonstrating that, just as human beings are part of the natural world, so too psychology is, or can be, an objective branch of natural science. In 1913, John Watson published a paper entitled ‘Psychology as the Behaviorist Views It’ in which he urged the replacement of mental states with behavioural contingencies as the proper objects of psychological study. He envisaged that the observation of animals, including humans, when placed in appropriately controlled experimental settings, would lead to the discovery of behavioural generalisations that would ultimately give way to explanations in terms of the physical and chemical processes involved in the functioning of muscles and nerves. By changing the focus of psychology from mental states that were observed by introspection to behavioural acts that were subject to supposedly more direct and objective observations, behaviourism sought to rid psychology of its unscientific and somewhat mysterious status. Of course, underlying the experimental emphasis within behaviourism’s methodology was the assumption that behaviour is fully determined and
that it is in principle possible, although in practice difficult, to discover regular causal 
associations between environmental conditions and behavioural acts. Further, the 
discovery of these regularities was foreseen as preliminary to both the prediction of 
behaviour in the presence of certain stimuli, and the control of behaviour through the 
manipulation of these stimuli.

In declared opposition to the behaviourist movement is the philosophy and practice of 
psychological hermeneutics. In its original context, hermeneutics studied the principles 
and methods that guide the interpretation of written texts that for various reasons were 
not open to an unambiguous first reading (Føllesdal, 2001). The purpose of interpreting 
these texts was to bring to light their meaning; to explicate what was unclear within 
them so that they could be understood. Interpretative practices have long since been 
seen to have application outside the area of written text, including the understanding of 
meaningful human actions (Gauld & Shotter, 1977).

While the behaviourists rejected psychological processes as objects of study because of 
their assumed irrelevance to causal analyses of behaviour, psychological 
hermeneuticists reject causal analyses because of their assumed irrelevance to 
understanding psychological processes, including meaningful action. Martin Packer 
(1985), for example, claims that human action has a semantic rather than a causal or 
logical structure, and that explanation of such action requires interpretation that takes 
into account the social and historical background in which it occurs in order for its 
meaning to be revealed.

Meaning, it seems, is the feature that most clearly distinguishes between the physical 
and the psychological characteristics of a given action. The bringing down of an 
auctioneer’s hammer, for example, is both a physical descent of a wooden object, and 
an indication that the item at hand has been sold. In studying human behaviour, one is 
typically interested in the meaning of actions rather than their physical specifications. 
Whether the auctioneer used a heavy or a light hammer is most probably of little 
concern to both the buyer and the seller. For them, what is important is that the price of 
the item has been fixed and that an agreement has been made. This is what the 
hammering means.
When an action is deemed to be meaningful in the way that it is claimed the auctioneer’s hammering is meaningful, it becomes open to a variety of interpretations. We may ask in what ways, and for whom, the action holds meaning. This is where the process of interpretation becomes important in psychology. The conventional meaning of the hammering, it may be agreed, is that the item has been sold. Nevertheless, there may be many additional meanings, perhaps personal ones, attached to that same action. For the seller of the item, the hammering that signifies its sale might mean a release from debt, but also the loss of a treasured heirloom. The buyer, on the other hand, might understand the hammering to mean an increase in his or her social status, and an associated increase in hours that must be spent at work. The explication of meaning, whether conventional or personal, requires an acquaintance with social customs and also personal beliefs, feelings and endeavours. This explication, hermeneuticists maintain, is the primary goal of psychology.

The emphases of the behaviouristic and hermeneutic approaches to the study of action, it follows, differ vastly. Behaviourists have as their primary aim a functional or causal analysis of behaviour in terms of environmental contingencies, that is, factors residing outside of the organism. The aim of a hermeneutic interpretation of action, on the other hand, is “… first and foremost the giving of an account that is sensible in the way it addresses current interests and concerns, not a search for timeless and ahistorical laws and formal structures” (Packer, 1985, p.1088). The hermeneutic approach to psychology assumes that it is only through identifying a person’s cultural setting and individual history that one can attempt to elucidate the meaning of his or her behaviour. This contrasts sharply with the timeless laws that apply to the interactions of physical objects; the cause and effect relations that persist regardless of cultural epoch and past experiences. Consequently, the methods employed by psychologists from the behaviouristic and hermeneutic schools are quite different. While both are concerned with behaviour, the behaviourist tries to explain it through the methods of the natural sciences, with particular emphasis on causal analyses, whilst the hermeneuticist tries to understand it via interpretation according to principles that have been developed independently of the tools with which non-meaningful phenomena are examined.

The hermeneuticists’ justification for espousing a unique methodology for the study of human thought and action centres on the view that to use the methods of the natural
sciences within psychology, as does behaviourism, would be scientistic, that is, would be a kind of play-acting at science when in fact scientific methods are inappropriate (Martin & Thompson, 1997). Behind the hermeneuticist’s claim that the methods of the natural sciences are inappropriate for psychology are two driving forces: firstly, there are methodological concerns that both the discovery and manipulation of psychological causal regularities are beyond human capabilities (Martin, 1993); and, secondly, there is the more pressing concern that natural science methods are inappropriate in psychology because, “What is meaningful has quite different modes of Being from the objects of those sciences” (Jaspers, 1963, p.355). As might be expected, behaviourists, like hermeneuticists, are guided, albeit to contrary conclusions, by methodological and ontological theses. To these I now turn.

I.iii Methodological issues

Initially, methodological concerns appear to have supplied the primary motivation for behaviourists to ignore psychological phenomena in causal analyses. Watson (1913) wrote that, “One can assume either the presence or the absence of consciousness anywhere in the phylogenetic scale without affecting the problems of behavior by one jot or one tittle; and without influencing in any way the mode of experimental attack upon them” (p.161). Watson’s approach does not deny the existence of consciousness, it simply ignores it as irrelevant to experimental methods.

B.F. Skinner (1953), too, makes claims that appear to serve similar practical interests, such as, “The practice of looking inside the organism for an explanation of behavior has tended to obscure the variables which are immediately available for a scientific analysis” (p.31). The accessibility of external events for functional analyses seems to suggest to Skinner that these are the events most appropriately focused upon in behavioural studies. Even if there does exist a causal chain from an environmental stimulus (the first link) to an inner state of the organism (the second link) to organismic behaviour (the third link), Skinner advises that “… we may avoid many tiresome and exhausting digressions by examining the third link as a function of the first. Valid information about the second link may throw light upon this relationship but can in no way alter it” (p.35). Once more, it is methodological ease that appears to be directing the behaviourist to ignore the causal relevance of psychological processes rather than uncertainty that such processes exist or are causally efficacious.
The behaviourists are not without company in employing methodological issues to promote their mode of investigation. One proponent of hermeneutic methods, Jack Martin (1993), makes a strong case concerning the doubtful prospects of ever discovering causal mechanisms with which therapists could potentially help their patients. He does not deny that such mechanisms might be at work, but argues that it is only in artificial contexts such as laboratories that causal regularities can be discovered, and, further, that the displacement of persons from their social, linguistic, cultural and historical contexts ultimately alters the psychological phenomena that the therapist wishes to explain. The conundrum is thus that it is only in real-world contexts that psychological phenomena can be observed whilst it is only in controlled artificial situations that causal regularities can be observed. Consequently, practical barriers appear to necessitate abandoning the search for psychological causal mechanisms, and likewise the goals of prediction and control which are so often paired with the quest to discover these mechanisms.

Methodological and epistemological difficulties in discovering and utilising psychological causal regularities will of course be of great importance to the practice of psychology insofar as it aims to alter what are considered to be negative thoughts, feeling and behaviours in persons seeking psychological counselling. However, the ease or difficulty with which one is able to apply the methods of psychology has no bearing on what is a vastly more important question, whether indeed psychological phenomena are caused and in their turn causally efficacious. As both Skinner (1953) and Martin (1993) agree, one’s inability to detect causal mechanisms is no guarantee that none are there to be found. If the hermeneuticist’s suspicion that causal mechanisms are absent from psychological processes were based wholly upon methodological considerations such as these, then it would be resting upon very unsure soil. Similarly, the fact of ‘ease of experimental methods’ would be a highly unsatisfactory reason for ignoring psychological processes in behavioural analyses, and, indeed, Skinner (1963) is critical of those who, for purely methodological reasons, ignore the role of thought processes in producing behaviour whilst assuming them to exist. In addition to their respective methodological concerns are the discrepant ontological assumptions that behaviourists and hermeneuticists have concerning the characteristics of mental phenomena, and it is these that are primarily responsible for directing the behaviourist towards, and the hermeneuticist away from, seeking causal regularities in human behaviour.
I.iv Ontological issues

The methodological discrepancy between the behaviourist and the hermeneuticist reflects the opinion of the former, that human action is a species of physical event for which one need develop no methods in addition to those of the natural sciences because humans form part of the system with which natural science is concerned, and of the latter, who maintains a physical-psychological dualism.

Skinner (1974) rejects the notion that one can maintain a dualism between psychological and physical events, whilst supposing that these interact causally, in the following way, “The puzzling question of how a physical event causes a mental event, which in turn causes a physical event, remains to be answered or dismissed as unanswerable…” (p.211). The assumed absence of physical dimensions in mental events is, he argues, a reason for rejecting them. A science of behaviour, as Skinner envisages it, “… must consider the place of private stimuli as physical things, and in doing so it provides an alternative account of mental life” (p.211). The alternative account mentioned here incorporates the facts of physiology with the environmental contingencies to which an organism is exposed, and provides a more complete account of behaviour than can knowledge of those contingencies alone. What appears is an explanation of behaviour in purely physical terms with mentalistic descriptions being disregarded as pre-scientific superstitions.

One immediate advantage of rejecting mentalistic accounts of behaviour in favour of purely physical accounts is that the latter are clearly amenable to causal analyses in a way that the former have not been. As Skinner (1953) suggests, “Prevailing philosophies of human nature recognize an internal “will” which has the power of interfering with causal relationships and which makes the prediction and control of behavior impossible” (p.7). No such interference occurs in a purely physical account of behaviour from which mentalistic states and events have been discarded as mere fancy.

The claims of dualism and self-determination that are inherent in traditional mentalistic accounts of behaviour therefore brought Skinner to the conclusion that the explanation of human behaviour requires knowledge only of environmental contingencies and, where possible, physiological facts. Consequently, mentalistic explanations of
behaviour were viewed as antithetical to causal explanations, and mental entities and events were deemed to be mere inventions, absent from the natural world.

Hermeneuticists also recognise the traditional dualism between physical and psychological events and the consequent impossibility of causal interaction between the two, however, rather than attempt to do away with the ‘mental realm’, they embrace the dualism and forsake the goal of psychological causal analysis in favour of understanding meaningful, rational, human action.

In their attempts to elucidate the meaning inherent in intentional actions, hermeneuticists aim to discover a rational configuration of beliefs and desires that renders the action intelligible. The process involves consideration of many types of influence external to the actor, such as cultural practices, social settings, the expectations of others and any other factor that could be useful in revealing how the particular action fits into a larger psychosocial context. It is in response to the demands of developing a rationally coherent picture of meaningful action that Karl Jaspers (1963) concludes that “… the understanding of meaning demands other methods than those of the natural sciences” (p.355). His argument proceeds from stating that natural science methods are inhibitory to the understanding of meaning to the suggestion of several principles by which meaning can be investigated. These principles are a mixture of ontological claims concerning the characters of meaningful phenomena, and suggested methods that are tailored to best interpreting them. Four of the six principles are included and elaborated on in what follows in order to demonstrate that psychological hermeneuticists view psychological phenomena as having differing modes of existence from physical phenomena.

**A. Empirical understanding is an interpretation**

“What is meaningful only has empirical reality in so far as it appears in perceivable facts. It is related to this that all *empirical* understanding is an *interpretation*” (p.355). Jaspers’ first principle for the elucidation of meaning indicates that meaning is manifested in the expressions, actions and creations of individuals, and that it is only through coming into contact with these manifestations that an observer can begin to understand the meaning of another. It follows that all understanding of another’s meaning is an interpretation of that meaning’s physical manifestations, and,
furthermore, that regardless of the degree of certainty one comes to feel when his or her interpretation appears to concur with all of the perceived facts, an alternative interpretation is always possible.

B. Understanding follows ‘the hermeneutic round’

“… we may only understand the particular from the whole but the whole may only be understood via the particular” (p.355). Following the ‘hermeneutic round’ or ‘hermeneutic circle’ refers to the to-and-fro process involved in interpreting a person’s actions. It is only in the context of the entire psychological make-up of a person that a component part, such as an individual action, can be understood. However, holistic understanding, too, is reliant upon interpretation of the psychological component parts, such as individual actions. Consequently, there is no optimal starting point for the interpretation of the meaning of a person’s actions, rather it is necessary to proceed to-and-fro between the individual components and the larger picture, each influencing how the other is to be understood. Furthermore, there is no objective endpoint at which the movement from part to whole and back again is no longer necessary. Interpretation is an ongoing process.

C. Opposites are equally meaningful

“Everything that is meaningful moves in opposites, and it is related to this that, methodologically, opposites are equally meaningful” (p.355). Jaspers relates that just as it is understandable that someone feeling weak and wretched may also feel spiteful and hateful towards better off others, it is equally understandable that the wretched person may feel lovingly towards others; opposite understandings are equally plausible. The lesson here is a methodological one which warns against assuming, a priori, one interpretation over its opposite without due consideration of all of the facts. From this principle one discerns the futility of attempting to predict future behaviour, and subsequently the implied rejection of a key behaviourist goal.

D. Unlimited interpretation

“The particular, whether an objective fact, an expression, intended content or act or indeed any single psychic phenomenon, loses meaning when isolated but gains meaning in context. It is related to this that all phenomena are open to unlimited interpretation and reinterpretation, just at the point where understanding stops” (p.356). This
principle asserts that the apparent establishment of a definite interpretation of meaningful phenomena is quickly overturned as an alternative interpretation presents itself. This suggests a fluid interchange between the process of understanding and the objects of understanding. Even when the objects of understanding remain unaltered, the understanding itself progresses and becomes refined, finding new meaning in what has previously been only partially understood. The infinite process of reinterpretation, Jaspers argues, precludes the use of the methods of the natural sciences, their criteria being inadequate for the revealing, or construing, of meaning. Instead, what is required in interpretation is the development of a coherent and complex although tentative account of what is observed. With the accretion of facts comes greater certainty of understanding, but this process does not lead to the elimination of alternative interpretations.

The principles outlined here for the guidance of the hermeneuticist are clear in suggesting that the objects with which psychology is concerned, human beings, are not to be adequately understood through the employment of traditional scientific approaches. While the hypotheses of the latter admit of falsification and subsequent rejection, incompatible interpretations concerning human actions need not suggest that one or the other must be discarded, so long as each is consistent with empirical observations. Consequently, multiple interpretations concerning any one action may coexist peacefully without fear of displacement. This situation is understood to be a product not of the complacency of those espousing rival interpretations, each content to uphold his or her own version, but rather of the inconclusive or indeterminate character of the objects of interpretation themselves. Underlying the hermeneutic method is the assumption that interpretation is not simply a passive means of understanding what exists independently of being studied, rather, that in the absence of interpretation all meaning ceases to be. This is to say that meaningful thought and action have no existence independently of interpretation, or that it is only through their being interpreted that thought and action become meaningful. Rather, then, than have an objective independent existence, psychological phenomena are presumed to have a subjective existence, one that is dependent upon third person interpretation.

The constitutive element of psychological theories, whereby what is studied exists because it is studied, is a feature that is wholly absent from the natural sciences. We do
not, for example, believe that our circulatory systems came into being only with their
discovery, rather, that they could be discovered precisely because they existed already.
Likewise, the movements of the planets proceed indifferently to the opinions, wishes
and assumptions of human observers. The existence of natural phenomena such as
circulatory and solar systems is well understood to be prior to, and independent of, any
investigations concerning them. It is only when we come to the human sciences, such as
aesthetics, ethics, and psychology, that the independent existence of what is studied is
either questioned or outrightly denied. Is a painting beautiful if there is no-one who
judges it to be so? Is an action wrong if there is no-one who condemns it? These
questions concern the conditions of existence of aesthetic and moral properties. The
interpretative approach taken in psychology raises like questions: Do individuals think
thoughts and perform actions independently of, and prior to, being interpreted as
thinking thoughts and performing actions? Do psychologists discover rational patterns
in human thinking and behaviour that need not have been discovered in order to have
existed?

As has been suggested, psychological hermeneuticists are implicitly committed to the
denial of the independence of rational behaviour from interpretative processes.¹
According to them, psychological investigation does not simply expose its subject
matter, it simultaneously creates it. It is this that primarily sets apart the rational mental
phenomena of psychology from the non-rational phenomena of the natural sciences, and
in this we see why both behaviourists and hermeneuticists have concluded that mental
and causal processes are incompatible.

Nevertheless, there remains the possibility that the behaviourists and the
hermeneuticists have been mistaken in their conclusions, and that mental causation is
indeed possible. Two theories of mental causation are to be examined in the present
thesis. The first of these is the well known account of mental causation championed by
Donald Davidson (1917-2003), that of ‘Anomalous Monism’. Davidson adopts the
interpretative approach to psychology. He accepts that human beings are rational and
that as rational beings their psychological processes are exempt from the causal

¹ As Maze (1983) notes, hermeneutics need not take this approach because the meanings of social
customs and rituals are objective, and are therefore discoverable facts concerning the societies within
which they are practiced. Nevertheless, as demonstrated, the dependence of rationality upon interpretive
processes is assumed by at least some hermeneuticists, and among their numbers is Donald Davidson,
with whom I am primarily concerned.
regularities governing the physical world. Nevertheless, he recognises that thought is causally related to the physical world, that perception, for example, is an effect of external stimuli impinging on the sense organs, and that human behaviour impacts on the physical environment. Further, far from adopting the sort of mental-physical dualism that might be supposed of one who argues for the exclusion of mental events from physical causal laws, Davidson’s theory is monistic – he believes in the identity of mental and physical events. The paradox of human rationality in a non-rational deterministic world is, according to Davidson, resolvable. Part I of this thesis examines the account of mental causation offered by Anomalous Monism.

The second theory of mental causation to be considered derives from the work of a lesser known philosopher, John Anderson (1893-1962), and is known as ‘Direct Realism’. Direct Realism differs from Anomalous Monism in a variety of ways. It rejects the interpretative approach to studying psychological processes, arguing that they, like all natural phenomena, exist independently of being contemplated and thus do not require interpretation in order to exist. Further, it rejects the notion of free choice whereby thought and behaviour occur independently of causal conditions. Direct Realism asserts that the rationality of human thought, where it obtains, reflects the logical structure of the world rather than being an inherent characteristic of thought itself. The paradox that results from viewing humans as set aside from the rest of the world by virtue of their rationality does not arise for the Direct Realist, who considers mental phenomena to exist in space and time, that is, to be fully incorporated parts of the natural world. Acceptance of the Direct Realist account of mental causation does, however, require that one relinquish the oft-cherished conviction that human behaviour is inherently rational and purposeful. The account of mental causation put forward by Direct Realism is examined and developed in Part II of this thesis.
PART I: DAVIDSON’S THEORY OF ANOMALOUS MONISM

Section 1: Establishing the Paradox

In 1963, Davidson wrote a paper entitled ‘Actions, Reasons, and Causes’ which, as its title suggests, was aimed at defending the notion that agents perform actions because they have reasons to do so. Davidson viewed his paper as a defence of commonsense, understood literally. It was the folk psychology view that people behave in ways that are intelligible in light of their goals that Davidson wanted to re-establish. Of relevance to Davidson’s enterprise has been the prevailing opinion that to rationalise or understand an action in terms of an agent’s reasons is incompatible with a causal explanation of the action in terms of those reasons. It was with this opinion in mind that Davidson set out to demonstrate that providing a rational account of action is not only compatible with a causal account, but also that it requires a causal account. Section 1.1 reviews Davidson’s arguments for human rationality, including the claim that action is rational only if it is caused by an agent’s reasons.

The causal relevance of mental events as advocated by Davidson is to be understood as extending beyond situations where both the cause and the effect are psychological. Physical events, Davidson argues, cause mental events, and mental events cause physical events. A natural conclusion to be drawn from the causal interaction of physical and mental events is that the two are lawfully related. This conclusion, however, Davidson strongly denies. Far from conceding that mental events form regular connections with physical events, Davidson argues that they are anomalous. The anomalous character of mental events, according to Davidson, means that they are lawfully related to neither other mental events, nor physical events. Strict deterministic laws, he asserts, are the province of physics and should not be sought in psychology. Davidson’s arguments for the anomalous character of mental events are reviewed in Section 1.2.

Combined with his claim that all instances of causation involve strict laws, Davidson’s view that mental events are both causally efficacious and free from strict laws results in a paradox. The nature of the paradox is delineated in Section 1.3.
1.1 Davidson and rationality

According to Davidson, there is something special about human beings that separates them from machines, plants, and, indeed, other animals. Humans are rational. The uniqueness of human rationality does not suggest that machines, plants and non-human animals are irrational, rather, they are non-rational. Considerations of rationality are irrelevant to the functioning of all that are non-human. By contrast, rationality, Davidson claims, is so essential to personhood as partially to define it.

Whence comes rationality? Why does it constrain, of all things in the universe, only human beings? According to Davidson (1982a), rationality is a social trait possessed only by those with language, and thus applies, as the only users of language, exclusively to humans. Those such as infants who are not yet language users, according to this criterion, are not rational (although they are likely to become so in the future). The defining feature of rational animals, Davidson argues, is the possession of propositional attitudes, of which he takes belief to be the most basic. The importance of language within Davidson’s theory derives from his view that in order to have beliefs, one must also have language. Language is therefore necessary for propositional attitudes, just as propositional attitudes are necessary for language. The two, Davidson (1999) claims, being necessary for each other, emerge together. It follows that all and only language users have propositional attitudes, and, the world being as it is, humans are the only creatures to possess both. Defined by the possession of propositional attitudes, the domain of rationality is therefore restricted to language using humans, or, as Davidson refers to this group, ‘people’.

It will be noticed that Davidson’s argument for human rationality is comprised of two claims, namely: that in order to have propositional attitudes one must have language; and, that the possession of propositional attitudes is sufficient for being rational.

The first claim, concerning the necessity of language for propositional attitudes, is composed of contentious elements such as that in order to have a belief, one needs the concept of belief; that in order to have a concept of belief, one needs also the concept of objective truth; and that one’s sense of objectivity arises through linguistic interaction
with other creatures (1975, 1988, 1989a, 1996, 1999). To explore these elements of the argument for the necessity of language for belief, while of interest in its own right, would be to divert the focus from Davidson’s larger goal, which is not to specify who or what is rational, but of those who are rational, how causal interaction with the non-rational world proceeds. For the current purpose, whether or not non-human animals can also be rational is but a side issue. Further, Davidson admits that he has not shown that language is necessary for a sense of objectivity, only that it suffices, and that his conviction of its necessity results because he can imagine no alternative way in which to achieve a sense of objectivity. It is thus inconclusive, both because of the contentious nature of the elements of the argument, and because even if the path from language to belief is unobstructed there may be alternative routes, whether language is at all necessary for having propositional attitudes. Rather than pursue these issues, a more expedient approach to understanding Davidson’s project is to accept that people do have beliefs, whether or not these depend on their also having language, and to discover why having beliefs and other propositional attitudes necessitates that one is rational.

The second claim, that possession of propositional attitudes is a sufficient condition for rationality, is central to Davidson’s thesis. Mental phenomena, or, as Davidson often refers to them, ‘the mental’, are typically divided into two basic sorts: bodily sensations and propositional attitudes. Davidson’s interest is confined to the latter. By ‘proposition’ is meant a situation such as ‘the curry is hot’ and by ‘attitude’ is meant such things as belief, hope, doubt, fear, surprise and so forth. Of course, each of the different attitudes may be paired with the one proposition, such as when someone both believes and is surprised that the curry is hot, while another doubts that the curry is hot. The distinguishing mark of the mental phenomena in which Davidson is interested is that they have content, are about something, or, as Brentano phrased it, are ‘intentional’.

While most would agree terminologically that what attitudes concern are ‘propositions’, there are various views as to the status of propositions themselves. Davidson (1989b) rejects the view that propositions are mind-dependent abstract objects that are known independently of the senses, arguing instead that the kinds of objects to which believers are related are things that can be pointed at or described. Such objects, Davidson claims, are not made subjective or ‘inner’ by being known, rather, they are objects and events.

---

2 In Part I, all date-references appearing in brackets without a name refer to Davidson’s publications.
that exist externally to the believer. Davidson’s ‘externalism’ will be examined more closely in Section 1.2.2.1.

Having indicated what propositional attitudes are, we come to Davidson’s (1982a) claim that “… to be a rational animal is just to have propositional attitudes, no matter how confused, contradictory, absurd, unjustified or erroneous those attitudes may be” (p.318). To what, then, does rationality amount?

As expounded by Simon Evnine (1991), Davidson’s sense of rationality is one in which a person’s system of attitudes is internally consistent. Internal consistency comprises two basic interacting elements: the holism and normativity of propositional attitudes. By ‘holism’, Davidson means that the propositional attitudes exist in a dense interrelated network which functions to give each individual attitude content and identity. This occurs partly through the relationships that each attitude has with the other attitudes. Davidson (1999) writes that, “Within any one dimension of mentality, such as belief, it seems clear that it is impossible to take an atomistic approach, because it is impossible to make sense of the idea of having only one or two beliefs. Beliefs do not come one at a time; what identifies a belief and makes it the belief that it is is the relationship (among other factors) to other beliefs” (pp.7-8).

To hold a belief, Davidson insists, requires the holding of many beliefs, which in turn means that one has either many beliefs or no beliefs at all. Importantly, the content of a belief is determined, in part, by the beliefs to which it is related. Consider, for example, the belief that ‘the striking of the match caused it to ignite’. While there is certainly no fixed set of accompanying beliefs, all of which are required if one is to have this thought, it is impossible that this thought could occur in complete isolation from all other thoughts. Rather, what is required is a large but unspecifiable network of further simultaneously held beliefs that fall into two basic categories, general and particular.

General beliefs, Davidson (1982a) says, must be empirically plausible, suggesting that they are beliefs about certain types of objects and events, of which the particular thing believed in is an instance, interacting in ways known to be possible. In the present example, one general belief held concerns the efficacy of friction in producing fire. This belief, too, will be connected to other beliefs, perhaps such as the belief that fire is hot,
and the belief that rough surfaces result in more friction when rubbed together than do smooth surfaces. These sorts of beliefs are general in the sense that they are about types of objects and events.

Alongside of general beliefs are particular beliefs. The thought being considered presently, that ‘the striking of the match caused it to ignite’, does not mean that in general the striking of matches causes them to ignite, but that the striking of a certain match, at a certain time, caused it to ignite. This is where particular beliefs have a role to play. Particular beliefs are ones such as that ‘the match was in fact struck’. The necessity of particular beliefs of this kind is obvious. If one did not believe that the match was struck, then the more complicated belief, that ‘the striking of the match caused it to ignite’, could not be held, because one of its elements would be absent. Of course, particular beliefs, while specific to a time and place, are still concerned with general sorts of things such as matches and ignitions, and belief in these requires that many more general beliefs be held. The identity of the original belief, that ‘the striking of the match caused it to ignite’, is therefore reliant, if Davidson is correct, on its relations to many other beliefs of both a particular and a general character. This is what he means by asserting that “… the intrinsically holistic character of the propositional attitudes makes the distinction between having any and having none dramatic” (1982a, p.318)

Tied in with the content giving relations that beliefs have with each other is their logical relations. It is here that the second element comprising rationality, normativity, joins with the first element, holism. Normative principles are simply the laws of logic applied to propositional attitudes. Just as the law of non-contradiction states that something cannot simultaneously be $P$ and $\neg P$ (e.g., a flower cannot simultaneously be red and non-red), normative principles state that one cannot simultaneously believe $p$ and $\neg p$. Similarly for the other laws of logic. The implication of the applicability of normative principles to the network of propositional attitudes is that the network exists as a logically coherent whole. Further, this condition, that beliefs and other attitudes be consistent with each other, is seen by Davidson (1999) to be a necessary one if any belief is to possess clear content.
Importantly, the coherence of the attitudes with each other is not a fortuitous outcome of the way in which they develop, nor of actively challenging one’s attitudes and discarding incompatible misfits, rather, the coherence results because the very identity of an attitude is a function of its place within the network, “Since the identity of a thought cannot be divorced from its place in the logical network of other thoughts, it cannot be relocated in the network without becoming a different thought” (1982a, p.321). The placement of a belief such as that ‘the striking of the match caused it to ignite’ within the network of attitudes relies not only upon the presence of the general and particular beliefs considered above, but also on the collective adherence of the beliefs to the normative principles. For example, if it were possible for one to believe simultaneously both that the match was struck and that the match was not struck (i.e., to believe in a contradiction), then there would be no clear content to the belief that ‘the striking of the match caused it to light’. Neither would there be a clear content to either of the contradictory beliefs. This is because the contents of beliefs in a holistic system are interdependent, each informing the others.

While Davidson most often uses beliefs to demonstrate the holistic and normativistic character of thought, rationality is deemed to characterise the entire network of propositional attitudes. Holistic and normativistic interdependence thus features in the relations between beliefs, doubts, surprises, fears and so forth, and also in the relations between each of these and evaluative attitudes such as desires. Each, Davidson (1999) claims, is necessary for the others, “There are, as I have argued, no beliefs without many related beliefs, no beliefs without desires, no desires without beliefs, no intentions without both beliefs and desires” (p.10). Indeed, due to its conceptual ties to the propositional attitudes, Davidson also regards behaviour or ‘intentional action’ as belonging to the mental, and hence rational, realm.

Intentional actions are those that are caused by what Davidson (1963) calls a ‘primary reason’. Primary reasons are comprised of a pro-attitude for actions of a certain kind and a belief that an action is of that kind. Contained within the pro-attitude for actions of a certain kind is the conviction that these kinds of actions are apt for bringing about a desirable consequence. As such, the types of actions towards which one may have a pro-attitude can be broad, as long as they each tend to the desired end.
Davidson (1963) argues that by performing intentional actions, the characteristic rationality of our propositional attitudes extends to our behaviour, “In the light of a primary reason, an action is revealed as coherent with certain traits, long- or short-termed, characteristic or not, of the agent, and the agent is shown in his role of rational animal” (p.678). These traits, of course, refer to the beliefs, desires and other propositional attitudes possessed by the agent. The attitudes work holistically and in accordance with the normative principles in bringing about an action that is consistent with them, whereby the action is seen to be rational in the light of the attitudes.

For example, if I desire to fetch my socks which I falsely believe to be in my drawer in my bedroom, then by seeking my socks in the drawer I am acting consistently with my desires and my beliefs, that is, my action is rational in the light of my attitudes. An irrational action would have been my seeking my socks on the clothes line, even if that were where they were, whilst believing them to be in the drawer. The simplicity of this example should not mislead one to think that only minimal beliefs and desires are involved in causing actions. The attitudes act holistically here as always. Therefore, part of the rationality of seeking my socks in my drawer involves compliance with other desires and beliefs, and the weighing of their claims. It was not mentioned that I believed my drawer to be free from a red bellied black snake, however, had I believed the opposite, then the primary reason that caused me to seek my socks, while still present, would have been overridden. Intentional action is therefore consistent not only with the primary reason that causes it, but also with the rest of the network of attitudes that allows it to happen.

The sort of account given here of my action of seeking my socks in my bedroom drawer, based upon my beliefs and desires, Davidson (1963) refers to as a ‘rationalisation’. Rationalisations, he contends, have a dual function. The first of these is to provide an interpretation that reveals the connections between propositional attitudes and actions. He writes, “When we ask why someone acted as he did, we want to be provided with an interpretation. His behaviour seems strange, alien, outré, pointless, out of character, disconnected; or perhaps we cannot even recognize an action in it. When we learn his reason, we have an interpretation, a new description of what he did, which fits it into a familiar picture. The picture includes some of the agent’s beliefs and attitudes; perhaps also goals, ends, principles, general character traits, virtues, or
vices” (p.679). In describing an action in terms of the primary reason that led to it, one shows that action to be a complementary part of a bigger picture. This, then, is the first function of a rationalisation.

The second function of a rationalisation, according to Davidson (1963), is to reveal the causal relation between the primary reason and the action. In urging that rationalisation amounts to causal explanation, Davidson is attempting to bridge the understanding-explanation divide that has traditionally been understood to distinguish psychological interpretation from causal explanation. It is in reference to this that he writes, “Noting that nonteleological causal explanations do not display the element of justification provided by reasons, some philosophers have concluded that the concept of cause that applies elsewhere cannot apply to the relation between reasons and actions, and that the pattern of justification provides, in the cause [sic] of reasons, the required explanation. But suppose we grant that reasons alone justify actions in the course of explaining them; it does not follow that the explanation is not also – and necessarily – causal” (p.678).

In revealing the causal connection between primary reasons and actions, something more occurs in rationalisation than simply developing a coherent psychological picture. It is not sufficient for an action to be deemed rational that an agent possesses beliefs and desires justifying the action. What is further necessary for an action to be understood as rational is that it was caused by the appropriate reasons. If the rationalisation of an action means only that each of the relevant beliefs, desires and actions is present, then “… something essential has certainly been left out, for a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason” (1963, p.679).

A causal connection between reasons and actions is therefore necessary if actions are to be rational in the light of an agent’s reasons. However, there is a further condition upon the rationality of actions: they must be caused by the reasons in an appropriate way. The necessity of an appropriate causal connection between reasons and actions in rationalisations is demonstrated by Davidson through the notion of ‘deviant causal chains’. Consider the following, “… when we offer the fact of the desire and belief in explanation, we imply not only that the agent had the desire and belief, but that they
were efficacious in producing the action. Here we must say, I think, that causality is involved, i.e., that the desire and belief were causal conditions of the action. Even this is not sufficient, however. For suppose, contrary to the legend, that Oedipus, for some dark oedipal reason, was hurrying along the road intent on killing his father, and, finding a surly old man blocking his way, killed him so he could (as he thought) get on with the main job. Then not only did Oedipus want to kill his father, and actually kill him, but his desire caused him to kill his father. Yet we could not say that in killing the old man he intentionally killed his father, nor that his reason in killing the old man was to kill his father” (1974, p.81).

Despite having reasons for wanting to kill his father that would rationalise the killing of the old man had Oedipus known who he was, Oedipus’ ignorance of his relationship to the old man debars the interpretation of the killing in terms of the son’s reasons for wishing his father dead. The use of deviant causal chains such as these demonstrates that a causal chain between reasons and actions, while necessary for the rational interpretation of human behaviour, suffices only if it proceeds in an appropriate way. Davidson (1973a) writes that, “The point is that not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route… Beliefs and desires that would rationalize an action if they caused it in the right way – through a course of practical reasoning, as we might try saying – may cause it in other ways. If so, the action was not performed with the intention that we could have read off from the attitudes that caused it” (pp.78-79).

The upshot of all this is that, without an appropriate causal connection, the notion that an action is rational in the light of an agent’s reasons for performing the action is vacuous. The traditional hermeneutic view that interpretation of human behaviour primarily involves drawing together reasons and actions, whilst eschewing causality, is shown to be ill formed; there can be no separation of the understanding of human action through rational interpretation from causal explanation of that same action. It is because they are the effects of primary reasons that actions partake of human rationality.

It should now be clear as to why Davidson equates being rational with being a holder of propositional attitudes. Propositional attitudes are, he argues, constitutively rational in
the sense that they function as logically consistent networks. Anyone possessing a single attitude, it follows, has a whole host of attitudes. Further, the identities of the attitudes are interdependent, each being logically compatible with all of the others. From the assumptions of holism and normativity the conclusion of rationality follows, “To have a single propositional attitude is to have a largely correct logic, in the sense of having a pattern of beliefs that logically cohere. This is one reason why to have propositional attitudes is to be a rational creature” (1982a, p.321).

Moreover, just as the possession of propositional attitudes and the performance of intentional actions are conceived by Davidson to be sufficient criteria for rationality, they are likewise considered to be necessary. Davidson’s conviction of the interdependence of rationality and mentality leads to his claim that, “To the extent that we fail to discover a coherent and plausible pattern in the attitudes and actions of others we simply forego the chance of treating them as persons” (1970a, pp.221-222). That is to say, one either thinks and acts rationally, or not at all.

Just as all that think and act are rational, all that do not think and act are non-rational. There is, therefore, something about the mental processes of people that separates them from the processes of the rest of the natural world. One important implication of the rationality of mental phenomena and the non-rationality of non-mental phenomena, according to Davidson, is that there cannot be lawful connections, including causal regularities, between the two. Davidson’s reasons for espousing this implication are considered in the following section.

### 1.2 Davidson and mental anomalism

Thus far, the constitutively rational character of the network of propositional attitudes has provided the focus from which Davidson’s psychological theory is to be understood. The time has come, however, to consider a second feature of the theory that is of comparable importance, the anomalous character of psychological phenomena. The two, it will be seen, are not unrelated.

According to Davidson’s principle of ‘the Anomalism of the Mental’, “… there are no strict deterministic laws on the basis of which mental events can be predicted and explained” (1970a, p.208). Just as rationality is understood to be a constitutive feature
of thought and action, so too the anomalous character of mental events is taken by Davidson to be an ‘undeniable fact’, one that is a natural consequence of the character of mental phenomena.

Mental anomaly is comprised of two related elements. Firstly, there are no lawful regularities by which classes of mental events may be definitionally reduced to classes of physical events. Secondly, there are no strict deterministic generalisations such that certain mental events are invariably causally linked with other events, whether mental or physical. Consequently, Davidson argues, there is a necessary absence in psychology of strict deterministic laws of both the mental-mental (psychological) and mental-physical (psychophysical) type. This situation contrasts with that of the physical sciences, the events of which proceed according to strict laws in such a way as to allow for both their accurate prediction and explanation.

The distinction between the strict laws of physics, and the mere generalisations of psychology, is repeatedly emphasised by Davidson, and is of the utmost importance, he believes, for deriving a monistic theory of mind, that is, one in which individual mental events are identical with individual physical events. As Davidson himself admits, the suggestion that psychophysical monism follows from psychophysical lawlessness is a surprising one, and for the meantime it will be put aside (to be taken up again in Section 2.4). My present concern is, firstly, to make clear on what grounds Davidson distinguishes the strict deterministic laws of physics from the rough generalisations of psychology, and, secondly, to draw out the features of psychological phenomena that differentiate them from physical phenomena so as to make them unsuited for inclusion in strict laws, both definitionally and causally.

1.2.1 Strict laws and generalisations

In reply to his own question “… what should we count as a law?”, Davidson (1995a) claims that “Laws must be true universally quantified statements. They also must be lawlike: they must support counterfactuals, and be confirmed by their instances (these conditions are not independent). To qualify as strictly lawlike, they should contain no singular terms referring to particular objects, locations or times (strictly lawlike statements are symmetric with respect to time and location). Strictly lawlike statements do not contain open-ended phrases like “other things being equal”, or “under normal
conditions”. It must be admitted that such phrases are, tacitly or explicitly, part of the content of many legitimate laws; thus many laws are not strict, including the laws peculiar to such sciences as geology, biology, economics, sociology and psychology” (pp.265-266).

According to Davidson, laws are linguistic entities that make possible the explanation of past events as well as the prediction of future events. Laws are general in the sense that they do not refer to particular objects or events directly (for example by name), rather, they apply to relevant types of objects and events, whether presently existing or simply hypothesised about. The lawlikeness of laws, Davidson (1970a) believes, is a matter of degree, such that some laws are strict while others are less so, or not at all. To count as strictly lawlike, a law must in principle be free from ceteris paribus clauses, formulated in a ‘closed’ or homonomic vocabulary, and a priori knowable. These three conditions, Davidson claims, are satisfied by the laws found in physics, but not by those found in the ‘special sciences’, being most notably absent from psychology. A discrepancy exists, therefore, between the strict laws of the physical sciences and the so-called ‘non-strict laws’ or mere generalisations that are found in psychology. The elements underlying this discrepancy will be considered in turn.

Freedom from ceteris paribus clauses within a law is an ideal which, Davidson argues, is most closely approximated by the laws of physics. Knowledge of a law that is completely free from any sort of qualification of the ‘all things being equal’ sort brings with it the possibility of perfect predictions given certain starting conditions. Of the ideal law, Davidson (1970a) claims that, “…there is a theoretical asymptote of perfect coherence with all the evidence, perfect predictability (under the terms of the system), total explanation (again under the terms of the system)” (p.219). Davidson is willing, however, to countenance that the ideal law he envisages may be elusive even in physics, and, that rather than the strictest laws yielding perfect predictions, something short of this, such as probabilities, might be the best that can be expected. Nevertheless, there are, Davidson (1963, 1974) stresses, important differences between the kinds of probabilistic statistical generalisations found in physics and those found in psychology.

One of these is that prediction, in physics, is possible because one can generally tell in advance of a physical event whether the appropriate conditions for that event’s
occurrence are satisfied. Further, if the conditions are not satisfied, then one can make appropriate allowances.

The situation, Davidson (1963, 1974) argues, is very different in psychology. Psychological generalisations are replete with generous escape clauses that make accurate prediction impossible. Following a piece of behaviour, it may be feasible to retrodict what beliefs and desires the person behaving had, and thus to explain why the action was performed, but prior to the performance of the action there is no way of moving from loose generalisations between beliefs, desires and actions to accurate predictions of behaviour. Any behavioural prediction, if found false, can always be excused by supposing that another belief or desire interacted with those that were used in making the prediction, and this process of post hoc patching of the original law can proceed without end. Davidson (1974) concludes that, “What is needed in the case of action, if we are to predict on the basis of desires and beliefs, is a quantitative calculus that brings all relevant beliefs and desires into the picture. There is no hope of refining the simple pattern of explanation on the basis of reasons into such a calculus” (p.81). The hope that such a calculus may, in time, be developed, Davidson (1963) labels ‘delusive’, suggesting that psychological generalisations are imprecise not just because of our present ignorance of relevant strict laws (a state which physics, too, was in in its youth), but because there are no formulable strict psychological laws.

The second element of discrepancy between physical and psychological generalisations concerns the number of conceptual vocabularies (such as mental, physical and moral) that it is necessary to draw from in order to formulate a law. This is important for Davidson because he believes that “… a law can hope to be precise, explicit, and as exceptionless as possible only if it draws its concepts from a comprehensive closed system” (1970a, p.219). Laws that derive from ‘closed systems’ refer to just one class of concepts (e.g., physical concepts), and are accordingly referred to by Davidson as ‘homonomic generalisations’. On the other hand, laws that do not derive from ‘closed systems’ necessarily refer to more than one class of concepts (e.g., physical and moral concepts), and these Davidson calls ‘heteronomic generalisations’.

While one may expect to find homonomic generalisations in the physical sciences, laws that require only words that come from the physical vocabulary, there is no
corresponding singular vocabulary with which to formulate psychological generalisations. The mental vocabulary is insufficient for this task because formulation of psychological generalisations typically requires mention of non-mental concepts, or, as Davidson (1970a) expresses it, “the mental does not… constitute a closed system. Too much happens to affect the mental that is not itself a systematic part of the mental” (p.224). Consequently, psychological generalisations must be heteronomic, that is, generalisations that require words from more than one conceptual vocabulary.

The advantage of homonomic generalisations, Davidson (1970a) suggests, is that they are those “…whose positive instances give us reason to believe the generalization itself could be improved by adding further provisos and conditions stated in the same general vocabulary as the original generalization” (p.219). That is to say, the sharpening of a homonomic generalisation into a strict law, while retaining the use of the original conceptual vocabulary, is feasible because the additions that turn the generalisation into a strict law derive from the same class of concepts, and are stated in the same conceptual vocabulary, as the original generalisation.

The case is different, however, with heteronomic generalisations. While it might be possible to formulate strict laws from what were originally heteronomic generalisations, this is done at the price of abandoning at least one of the conceptual vocabularies that the original heteronomic generalisation was stated in. Thus, heteronomic generalisations “… may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary” (1970a, p.219). The ‘different vocabulary’, of course, is the physical vocabulary, because only this suffices for homonomic generalisations. The point that Davidson would have us understand here is that stated in familiar psychological terms such as ‘belief’, ‘desire’, ‘behaviour’ and so on, no psychological generalisation can be made strict. If there is indeed a strict law at work, this is one that can be captured only by casting off the psychological descriptions of the events and objects involved, and replacing these with physical descriptions. In this vein Davidson (1963) claims that, “The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalizations must deal” (p.684). As rationalisations deal with mental concepts such as desire and belief, this statement amounts to the claim that even if desires and beliefs cause actions, the laws required by any instance of causality will not be formulated in
terms of the desires and beliefs so classified, rather, “… the classifications may even be neurological, chemical, or physical” (p.684).

Taken in conjunction with the assumption that only laws that derive from closed systems can be strict (a claim for which Davidson provides no argument), the heteronomic status of psychological generalisations means that they are not capable of being made strict unless the psychological vocabulary is abandoned, in which case the resulting law, while strict, will not be psychological.

The third element underlying the discrepancy between physical and psychological generalisations concerns what Davidson sees as the *a priori* nature of physical laws, and the absence of equivalent *a priori* psychological and especially psychophysical laws. The ruling of a statement lawlike, Davidson argues, occurs prior to, and independently of, the discovery of empirical evidence for that statement. He writes that, “Of course, to judge a statement lawlike or illegal is not to decide its truth outright; relative to the acceptance of a general statement on the basis of instances, ruling it lawlike must be *a priori*” (1970a, p.216). The essence of this claim is that laws are formulated independently of the observations that will show them to be true or false. That is to say, rather than *discover* laws through empirical observation, scientists *create* laws which are conceptually coherent, which connect objects and events in a logical manner, and only then attempt to observe instances of the laws in real world situations.

The *a priori* character of physical laws may be observed, Davidson (1970a) argues, in the practice of measuring length. He contends that the concept of length, and consequently the practice of length measurement, each depends on the law of transitivity (i.e., if \( a \) is longer than \( b \), and \( b \) is longer than \( c \), then \( a \) is longer than \( c \)). There is no making sense of the concept of length, Davidson claims, unless the law of transitivity holds. Without transitivity, it would be impossible to assign numbers to rank lengths, and the demands of a ratio scale of measurement could not be met. With this in mind, Davidson asks what one could say if it was *thought* that an intransitive triad had been observed (e.g., where \( a \) is longer than \( b \), \( b \) is longer than \( c \), and \( c \) is longer than \( a \)). He summarily rejects the possibility that one could simply count the law of transitivity false, because this alternative would likewise mean that the concept of length, too, must be disposed of. The solution, Davidson suggests, is that “It is better to say the whole set
of axioms, laws, or postulates for the measurement of length is partly constitutive of the idea of a system of macroscopic, rigid, physical objects” (p.221).

What is to be made of this solution? If the observation of an intransitive triad cannot show the irrelevance of transitivity to length, then nothing can; and this is Davidson’s point. The concept of length is conceptually knitted to the law of transitivity. The relevance of the law of transitivity to the concept of length is *a priori*, knowable independently of observation, and robust to any imaginable falsification through observation. The idea of rigid physical objects too, it seems, is so tied up with the laws of measurement as to be partly defined by them. Consequently, it is only through having a theory of physical objects, which necessarily involves the *a priori* acknowledgement of laws such as that of transitivity, that one can speak of an object as having a length at all.

The present ‘solution’ to the problem of delinquent observations is reminiscent of the holism of propositional attitudes that was considered in Section 1.1. Just as the content or meaning of a belief is derived in part from the relationships the belief stands in to other propositional attitudes, and just as these too are informed by the belief, so too the notions of ‘length’, ‘rigid physical object’ and ‘transitivity’ combine to give content to each other. It is thus conceptual pressures, not observations, that sustain concepts such as length and that likewise make physical laws *a priori* knowable. The *a priori* nature of laws relating to the measurement of length is furthermore taken by Davidson to extend to physical laws in general, “I suggest that the existence of lawlike statements in physical science depends upon the existence of constitutive (or synthetic a priori) laws like those of the measurement of length within the same conceptual domain” (1970a, p.221).

One consequence of the conceptual interdependence of the elements of physical systems is the resulting ability one has to formulate laws independently of observation, “Nomological statements bring together predicates that we know a priori are made for each other – know, that is, independently of knowing whether the evidence supports a connection between them” (1970a, p.218). In this way one can bring together the ideas of ‘rigid physical object’ and ‘transitivity’ because these concepts are ‘made for each other’, neither is meaningful without the other. There is no requirement, and indeed no
possibility, of firstly observing the lengths of objects and only then discovering if the
lengths stand in transitive relations to one another, as if this were an open question. If
the objects have lengths, then it is known independently of observation that they will
conform to the laws of transitivity.

Propositional attitudes and actions, too, have conceptual interdependence with each
other and in this sense their potential combinations are constrained, just as the concept
of length is constrained by the axioms of measurement. Davidson (1970a) holds that,
“Just as we cannot intelligibly assign a length to any object unless a comprehensive
type theory holds of objects of that sort, we cannot intelligibly attribute any propositional
attitude to an agent except within the framework of a viable theory of his beliefs,
desires, intentions, and decisions” (p.221). Be that as it may, the constraints of
rationality on propositional attitudes do not produce strict constitutive psychological
laws in the way that the conceptual pressures of the physical sciences produce strict
physical laws.

While it may be assumed that ‘anyone believing $b$ will not simultaneously believe not-
$b$’, there is no way of knowing, independently of the evidence, that ‘anyone with belief
$b$ and desire $d$ will perform action $a$’ in the way that one can know that if object $a$ is
longer than object $b$, which is longer than object $c$, then object $a$ is longer than object $c$.
The reason for this difference, according to Davidson, is not the more complex nature of
making a psychological than a physical prediction, but the tentative nature of
psychological interpretation. Rather than exist in an a priori manner similar to physical
laws, psychological interpretations are constantly evolving as new evidence comes to
light, and the development of more coherent interpretations is an ongoing process
(recall Jaspers’ ‘unlimited interpretation’, Section I.iv). It thus follows that while
rationality constrains the combinations of propositional attitudes that may be held
simultaneously, debarring certain combinations, it nevertheless leaves open the way for
a variety of logically compatible configurations of beliefs, desires and actions. As a
result, the formulation of a priori psychological laws is an impossibility.

Moreover, while there is conceptual interdependence within both the physical and the
mental domains, there is no conceptual interdependence between these domains. We
cannot, that is to say, bring together, independently of the evidence, mental and physical
characteristics into lawful combinations because “… mental and physical predicates are not made for one another” (1970a, p.218). The impossibility of connecting, independently of observation, a belief such as ‘it is lunch time’ with a particular physical state such as a neural configuration, rules out, therefore, the option of a priori knowable psychophysical connections. If strict laws must be a priori knowable, then the absence of a priori psychophysical laws means that, at best, there can only be psychophysical generalisations.

The three conditions for being a strict law as understood by Davidson, the absence of ceteris paribus clauses, homonomicity and a priorism, are each, therefore, satisfied by physical but not psychological generalisations. Moreover, as has been hinted, the absence of strict laws from psychology has as its basis the character of psychological phenomena, and is therefore a permanent feature of any generalisations involving these. The following section aims to examine more fully the particular features of psychological phenomena that Davidson takes as underlying their anomalous character.

1.2.2 Mental anomalism: why strict laws are absent from psychology

It is clear that Davidson believes that psychological generalisations differ in kind from physical generalisations, that strict laws and the predictions that can be derived from them belong singularly to physical theories, and that psychological generalisations are necessarily imprecise, non-predictive, and hence, unlawlike. What is not, at this point, so clear, is what accounts for this schism between the physical world at large and the beliefs, desires and actions of a small proportion of its inhabitants, human beings. Why should there be no possibility of refining psychological generalisations into strict laws, and why, in particular, can there be no lawful connections between mental and physical phenomena? It is with this last issue, the impossibility of psychophysical laws, that Davidson’s theory is most concerned. His general approach to demonstrating the impossibility of such laws is to take physics as an exemplar of scientific methods and laws, and then to identify ways in which the phenomena and practices of psychology are unsuited to these methods. As a consequence, Davidson concludes that the differences between psychological and physical phenomena render them unnamable to lawful connection, and hence the anomalous character of psychological phenomena is seen to be established. The identified features of psychological phenomena that differentiate
them from physical phenomena, and account for much of the claim that the former are anomalous, are three: externalism, dispositionality and rationality.

1.2.2.1 Externalism

One characteristic of mental states that does not have its equivalent in physical states is their necessary intentionality or ‘aboutness’. Mental states have objects. There are not just beliefs, but beliefs about things, there are not just desires, but desires for things. One of the strengths of Davidson’s theory of Anomalous Monism resides in his acknowledgement that the objects with which thought is concerned are not, as has been assumed previously, somehow existing ‘before the mind’ in that they consist of ‘appearances’, ‘sense data’ or ‘qualia’ (1987a, 1989b). Rather than employ these ‘inner’ objects of thought, Davidson adopts a form of psychological externalism whereby the content of thought is determined by worldly objects that are external to the thinker’s body. In other words, Davidson’s view of mental states is ‘externalist’ because he believes that the identity of such states is determined, in part, by external factors, the objects and events by which they were caused.

Nevertheless, of the utmost importance to Davidson’s theory is the requirement that mental states be identical with physical states. This aspect of his theory, monism, as has been mentioned already, will be returned to in Section 2.4. A sense of it, however, can be gained from the following characterisation of mental states, “I think such states are ‘inner’ in the sense of being identical with states of the body, and so identifiable without reference to objects or events outside the body; they are at the same time ‘non-individualistic’ in the sense that they can be, and usually are, identified in part by their causal relations to events and objects [sic] outside the subject whose states they are” (1987a, p.444).

This dual sense of identity, whereby mental states are both internal bodily states and ‘about’ external objects and events, is usefully illustrated by analogy with an example of externalism in the physical world, involving a simple case of sunburn. Davidson (1987a) says, “My sunburned skin may be indistinguishable from someone else’s skin that achieved its burn by other means... yet one of us is really sunburned and the other not” (pp.451-452). The purport of this quotation is that sunburn is not merely a state of the skin, rather, it is a state of the skin with a particular causal history. As such,
accurately identifying a case of sunburn requires more information than is provided simply by observing the skin in its burnt state. The further information required concerns the means by which the skin came to be in that state, that is, what caused the burn.

The case is similar, Davidson (1987a) suggests, with mental states. A belief that ‘this rose is red’, is typically caused by perceiving that the rose in question is red, that is, it is determined by an external object. This belief, it follows from monism, will have an attendant bodily state, but the bodily state on its own does not determine the content of the belief, indeed, “… people can be in all relevant physical respects identical while differing psychologically” (p.453). Just as the way in which a burn was caused determines whether the skin is sunburnt or not, so too, the way in which a mental state was caused determines the content of that state. People in physically indistinguishable bodily states need not, it follows, be in identical psychological states. The identity of a psychological state is determined, at least partly, by the worldly objects and events that caused it.

While there are instances of physical states, such as sunburn, that are at least in part defined by their causal history, and are thus ‘external’ in a similar sense to beliefs, no such concepts, Davidson (1995b) claims, belong to a completed physics. Presumably, a completed physics that was dealing with sunburnt skin would be interested purely in the present state of the skin, and indifferent to the means by which the burn was acquired. However, Davidson (1983) argues that the externalism found in psychology is not so easily dismissed, “… if I am right, we can’t in general first identify beliefs and meanings and then ask what caused them. The causality plays an indispensable role in determining the content of what we say and believe” (p.317). Any attempt to reduce psychological attitudes to their accompanying bodily states is, it seems, futile. Without a particular causal history, a bodily state is simply a bodily state. With a particular causal history, a bodily state may also be a psychological state. It is the fact of the external objects and events causing the bodily state that makes the difference.

Given, then, that externalism is an essential feature of psychological phenomena, a distinction appears between the kinds of generalisations found in physics, that is, those that deal with physical states in and of themselves, and the kinds of generalisations
found in psychology, that is, those that rely upon both physical (bodily) states and the 
causal history of the acquisition of those states. As was asserted above, there need not 
be tight connections between the bodily states that are of interest to both psychology 
and physics, and the causal histories which are the province of psychology – recall that 
two people can be in physically identical states whilst thinking different thoughts. It 
thus follows that in dealing only with physical states, including bodily states, physics is 
ill equipped to account for mental content, which is partly dependent on factors external 
to the body. Consequently, the strict laws of physics do not apply to psychological 
phenomena. Furthermore, if it is indeed possible that two people can be in identical 
bodily states and yet be thinking different thoughts, then a psychophysical law of the 
type ‘when physical condition P obtains so too does mental content M’ is impossible to 
formulate. This is the argument for mental anomalism from psychological externalism.

1.2.2.2 Dispositionality

A second reason for psychological phenomena being unsuited to strict laws is what 
Davidson (1996) refers to as their ‘irreducibly causal’ character, which is more 
generally known as ‘dispositionality’. Dispositions are typically understood as 
potentials for objects to behave in certain ways under certain conditions. For example, 
‘fragility’ is characterised as follows, “… fragile things tend to break when struck hard 
enough, other conditions being right…” (1963, p.683). Further familiar examples of the 
many dispositions are brittleness, solubility and elasticity.

Despite typical English usage, ‘fragility’ and other dispositional names do not refer to 
states of objects, rather they refer to the responses objects make, or would make, under 
specifiable circumstances. The class of all fragile things are those that do, or would, 
shatter when lightly struck. To be fragile, therefore, is to respond to impact in a 
particular way. This tells one nothing about the state of the object itself, only what it 
does when it is struck. Similarly with solubility. Soluble things are those that dissolve 
when they come into contact with water. Knowing that sugar is soluble means knowing 
only that if sugar is placed in water, it will dissolve. Nothing is known about the state of 
the sugar itself.

Mental concepts, Davidson (1996) contends, are dispositional, “Beliefs and desires are 
identified in part by the sorts of action they are prone to cause, given the right
conditions” (p.171). Just like fragility and solubility, it is suggested that beliefs and desires are best understood in terms of how they respond in certain circumstances, specifically, through the kinds of actions that they are prone to cause. There is an important distinction, however, between dispositionality in physical objects and the dispositionality of beliefs and desires. With physical objects such as sugar, it is in principle possible to find out what it is about the state of the object that makes it soluble. Were one to find, for example, that loosely packed molecules make an object dissolve when placed in water, then one would know something about the object itself, and this would explain why the object behaves as it does. As Davidson (1990a) writes, “If a substance is soluble, there is something about its composition – an unspecified something – that causes it to dissolve under the right conditions. Science can say what that something is, and so dispense with solubility as an explanatory concept” (p.25). Consequently, physical objects such as sugar are suitable for inclusion in strict laws because it can be discovered in what state the object must be, and in which circumstances the object must be placed, in order for the expected result to occur.

The situation is different, Davidson (1996) argues, with mental states. They are irrevocably dispositional. There is no looking at the structure of a belief and discovering what it is about the belief that makes it cause an action under certain circumstances because, according to Davidson, “… it is part of the concept of a belief or a desire that it tends to cause and so explain actions of certain sorts” (p.172). That is, beliefs cannot be understood independently of the actions that they cause. Whereas with physical objects it is possible to move from understanding their responses (what they do) under particular circumstances to understanding them themselves (what they are), with beliefs and desires there is no such second step. There is no understanding beliefs and desires independently of what they do.

Because of their irrevocable dispositionality, Davidson (1996) argues that mental concepts are not suitable for use in strict laws. This is because no disposition is suitable for inclusion in a strict, exceptionless law. The reasons are twofold. Firstly, it is impossible to specify perfectly when the circumstances are such that the object will always behave in the expected way (that is, dissolve, shatter, cause an action). Dispositions refer to what objects usually do, there can always be exceptions. Secondly, dispositional ‘explanations’ tend to obscure, not explain, why objects behave as they do.
This is because they are circular. To say that something is soluble is to say that it dissolves when placed in water. To explain why something dissolved when placed in water by referring to its solubility is simply to restate what needs to be explained.

If mental concepts are irrevocably dispositional, and if dispositional concepts are unsuited for inclusion in strict laws, then it follows that mental concepts are unsuited for inclusion in strict laws. This is the argument for mental anomalism from psychological dispositionality.

1.2.2.3 Rationality

The most important arguments towards the anomalous character of psychological phenomena centre upon the distinction that Davidson draws between rational and non-rational systems. Broadly speaking, Davidson’s argument is that considerations of rationality are essential to mental processes but irrelevant to physical processes. Consequently, there cannot be lawlike connections between mental and physical processes.

The network of propositional attitudes, as conceived by Davidson, functions holistically and in accordance with normative principles, and is therefore internally consistent (see Section 1.1). In recapitulation, propositional attitudes interact to give each other content, with the whole network contributing to the identity of any one attitude. This means that beliefs and other attitudes cannot be formed one by one independently of what is already believed, rather that each attitude must cohere with the entire network, holistically and normativistically.

Neither holism nor normativity, Davidson (1970a; 1974; 1995b) argues, has its counterpart in the physical world. Physical systems, including human brains, do not function according to normative principles, nor do they function holistically. Brain states do not stand in the kinds of logical relations to each other that mental states do. As physical systems, brains follow the laws of physics. They respond to electrical and chemical stimulation by which they are either excited into firing electrical impulses, or inhibited from doing so. It is nonsensical to ask whether a particular brain state brought about by electrical or chemical activity is logically consistent with the brain’s functioning as a whole. The kinds of considerations that brains are responsive to are
non-rational. Consequently, “There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence” (1970a, p.222).

The requisite rationality of mental states in Davidson’s theory, together with the irrelevance of rational constraints in the physical world, means that there is no way of relating or equating types of mental states with types of physical states. In this, one sees Davidson’s commitment to anti-reductionism. Taken individually, Davidson believes that mental states are identical with bodily states, that is, every particular belief, desire, or other propositional attitude is a physical state of the body. However, there can be no lawful connections between types of attitudes and types of bodily states. For example, my belief that ‘the earth is round’ and your belief that ‘the earth is round’ will both, according to the theory, be bodily (presumably brain) states, but they may well be brain states that are physically distinguishable from each other (in ways other than being discovered in separate bodies). All that Davidson’s monism requires is that, taken individually, each propositional attitude is identical with a physical state, not that a type of mental state be invariably accompanied by a type of physical state. Accordingly, this version of monism is an instance of ‘token’, as distinct from ‘type’, identity.

The significance for Anomalous Monism of the absence of lawful connections between types of mental states and types of physical states should not be underestimated. Physical states, such as brain states, can be brought into being by other physical states and events within the relevant locality. Consequently, if there were a lawful one to one connection between types of brain state and types of psychological state, and if brain states respond to non-rational causal conditions, then it follows that psychological states, too, could be brought about by non-rational considerations, which violates the very notions of both holism and normativity, and thus rationality.

Holism would no longer constrain the network of propositional attitudes because as each brain state arose, so too would the relevant psychological state. However, the
relevance here would be relative not to already existing psychological states, rather, simply to the individual brain state with which the newly formed psychological state corresponded. If this occurred, one could be brought to believe that the earth is round while having no beliefs whatsoever concerning shape, planets, or anything else. Normativity, too, would disappear because if it were possible for the brain states corresponding to the earth being round and the earth being non-round to co-occur, thus causing the related beliefs, then contradictory beliefs could be held. If by definition psychological states are necessarily rational, then lawful connections between types of brain states and types of psychological states are impossible, precisely because brain states are not constrained by rational considerations (see Evnine, 1991).

According to Davidson, then, the constitutive rationality of propositional attitudes means that psychological phenomena are subject to different constraints from the rest of the natural world. The laws of psychology, if any there be, are those of rationality – that is, internal consistency through holism and normativity. By contrast, the rest of the world is answerable to non-rational physical laws. Consequently, “We must conclude, I think, that nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal” (1970a, p.223). This is the argument for mental anomalism from the rational character of psychological phenomena.

1.3 Davidson’s paradox

The anomalous character of psychological phenomena and the view, accepted by Davidson, that all instances of causation involve strict laws, combine to suggest the impossibility of causal relations between mental and physical events. It is somewhat paradoxical, therefore, that Davidson maintains that the psychological phenomena that are by their very nature anomalous are also causally efficacious. Davidson is well aware of the paradox contained in his theory of mental causation, yet he refuses to revise any of the central principles contained therein, claiming instead that they are capable of reconciliation.

It is in his famous paper from 1970, Mental Events, that Davidson first clearly set out the principles which, when taken together, appear to form an inconsistent triad. These principles are:
1. The Principle of Causal Interaction;
2. The Principle of the Nomological Character of Causality; and
3. The Principle of the Anomalism of the Mental

The first principle refers to the fact that physical events (such as bodily movements) can have mental causes, and that mental events (such as perceiving) involve physical causes, that is, that mental and physical events interact causally. The second principle asserts that causality involves strict deterministic laws, and therefore that any singular instance of causality falls under a causal law. The third principle claims that strict deterministic laws are not applicable to mental events for the reasons presented in the previous section. Any two of these principles, taken together, appear to entail the falsity of the third. Davidson, however, aims to demonstrate that the three principles are consistent with each other, and in so doing prove that mental events are simultaneously anomalous and causally efficacious. The arguments employed by Davidson to this end are described in Section 2.
Section 2: Resolving the Paradox

Davidson’s purpose in advocating Anomalous Monism is to reconcile the anomalous character of mental events with their causal and identity relations with the non-mental world. His argument rests upon a distinction he draws between the general character of laws and descriptions of events, and the particular character of events existing independently of their descriptions. Vital as this distinction is to Davidson’s argument, it is oftentimes overlooked, as Davidson (1993a) himself complains. As a result, much of the criticism of Davidson’s work fails to address the genuine issues that he raises, and understandably has had little more effect on Davidson himself than in drawing from him repetition and clarification of his original claims. However, the distinction Davidson draws between the generality of what he sees as purely linguistic phenomena relating to events (such as descriptions and laws) and the particularity of the events themselves, cries out for examination both because of its centrality to the resolution of his paradox, and because of the implications it has for the ontological status of mental properties.

Section 2 aims simply to outline Davidson’s assumptions and reasoning which together suggest to him that there is nothing paradoxical in maintaining simultaneously the principles of mental-physical causal interaction, causal nomologicality and mental anomaly. Evaluation of these arguments will be deferred until Section 3. The argument for the consistency of the three principles was originally formulated in *Mental Events*, and has since undergone very little change. The argument proceeds as follows,

It should now be evident how anomalous monism reconciles the three original principles. Causality and identity are relations between individual events no matter how described. But laws are linguistic; and so events can instantiate laws, and hence be explained or predicted in the light of laws, only as those events are described in one or another way. The principle of causal interaction deals with events in extension and is therefore blind to the mental-physical dichotomy. The principle of the anomalism of the mental concerns events described as mental, for events are mental only as described. The principle of the nomological character of causality must be read carefully: it says that when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law (1970a, p.215).
The essential claims within Davidson’s argument are the following:

1. Events are individuals;
2. Events are amenable to different sorts of ‘descriptions’;
3. Causal relations occur indifferently to the descriptions of the events involved; and
4. Causal laws are linguistic and therefore rely upon the types of descriptions of the events involved in causal relations.

Claims 1 and 2 form the subject matters of Sections 2.1 and 2.2 respectively, while claims 3 and 4 are addressed together in Section 2.3. Section 2.4 delineates Davidson’s argument for mental-physical event identity, that is, psychophysical monism. In support of the claims differentiating individualistic events from their universalistic descriptions, Davidson further argues that the concept of supervenience that applies in ethics is similarly applicable in psychology. The use to which Davidson puts supervenience is considered in Section 2.5, and the elements of the argument for solving the paradox of Anomalous Monism are brought together in Section 2.6.

2.1 Characterising events

2.1.1 Events as a fundamental ontological category

The potential success of Davidson’s approach to resolving his paradox relies crucially upon establishing that events are individuals, or, as he alternatively refers to them, ‘particulars’. In order, however, to argue that events are particulars, Davidson deems it necessary to demonstrate firstly that events, like objects, are things or entities, that is, that they exist. He refers to this task as a “… defence of events as constituting a fundamental ontological category” (1969, p.180).

Davidson’s stated aim in arguing for an ontology of events often takes the form of ‘making sense of our ordinary talk’ (1967a, 1969; 1970b; 1977; 1985a; 1985b, 1993b). This is because he believes that metaphysics is implicit in language, and, therefore, that by analysing our language we may discover facts about the things to which our linguistic devices refer. Thus, Davidson (1970b) poses the question, “Are there, in addition to pebbles and stars, movements, births, landslides, and explosions?”, and responds, “Our language encourages us in the thought that there are, by supplying not
only appropriate singular terms, but the full apparatus of definite and indefinite articles, sortal predicates, counting, quantification, and identity-statements; all the machinery, it seems, of reference. If we take this grammar literally, if we accept these expressions and sentences as having the logical form they appear to have, then we are committed to an ontology of events as unrepeatable particulars (‘concrete individuals’)" (p.181). Davidson’s procedure here is one of identifying aspects of language that would make sense only if there existed events to be referred to, and then concluding that these events must exist, “In short, I propose to legitimize our intuition that events are true particulars by recognizing explicit reference to them, or quantification over them, in much of our ordinary talk” (1969, p.166).

Davidson (1970b) also warns, however, against drawing ontological conclusions from what is suggested by language, and gives what is, in the present context, a more relevant reason for his view that events exist and that they exist as particulars, “A further need for events springs from the fact that the most perspicuous forms of the identity theory of mind require that we identify mental events with certain physiological events… I do not believe we can give a cogent account of action, of explanation, of causality, or of the relation between the mental and the physical, unless we accept events as individuals” (1969, p.165).

So what, then, are events? Davidson (1969) states, “We very often describe and identify events in terms of the objects to which they are in one way or another related” (p.173) and, further, “It even seems likely to me that the concept of an event depends in every case on the idea of a change in a substance…” (p.174). Here ‘object’ and ‘substance’ appear to be used interchangeably. Elsewhere, however, Davidson (1985b) suggests that events might not always require objects, as when they are actions people perform such as singing and running, “‘Gregory sang a song’, or ‘Peter ran to Canterbury’ seem, on the other hand, always to require an agent, but not to require an object. One may run without running anywhere, and sing without singing anything. But there is no singing without a singer. If this is right, we must think of singing as always represented by a predicate of an agent and an event; the event may then be characterized at will as having an object, or not” (p.232).
The feature common to both characterisations of events seems to be that of *activity*. An event is either a *change* in an object or an *action* performed by a person. The claims of activity are clearly discernible in Davidson’s (1969) attempts to discover when what appear to be two events, $x$ and $y$, are identical (that is, one). In this paper, Davidson considers and rejects the following options: Firstly, events are identical if they are changes in the same substance; secondly, events are identical if they are in the same place; thirdly, events are identical if they consume identical stretches of time; and fourthly, events are identical if they occupy exactly the same time and same place. The criterion that Davidson, at that time, settled on for discerning whether events are identical, is, “…events are identical if and only if they have exactly the same causes and effects… Not only are these the features that often interest us about events, but they are features guaranteed to individuate them in the sense not only of telling them apart but also of telling them together” (p.179).

What distinguishes the causal criterion from the others considered is that it is necessarily involved with activity. A cause brings something, the effect, into being. As Davidson (1969) remarks, “Events have a unique position in the framework of causal relations between events in somewhat the way objects have a unique position in the spatial framework of objects” (p.179). Nevertheless, Davidson (1985a) later abandons this position, accepting instead that “… events, like physical objects, are identical if they occupy the same places at the same times” (p.175). What is not abandoned, however, is acknowledgement of the fact that events involve activity, which is what allows Davidson to maintain that even if place and time together provide the criteria for individuating both objects and events, events can still be distinguished from objects, “… it may be, for example, that events *occur* at a time in a place while objects *occupy* places at times” (p.176).

Regardless of whether events are understood to be changes in objects or actions performed, Davidson’s main objective is to insist that rather than an asymmetrical dependence of events upon objects whereby events are somehow dispensable in a way in which objects are not, there exists instead a conceptual dependence of both objects on events, and events on objects. He writes that, “Substances owe their special importance in the enterprise of identification to the fact that they survive through time. But the idea of survival is inseparable from the idea of surviving certain sorts of change… neither
the category of substance nor the category of change is conceivable apart from the other” (1969, p.175). The purport of this and similar arguments is that events and objects are on an equal ontological footing.

2.1.2 Events as particulars
Having established in the above manner that events exist, Davidson requires to demonstrate that events are individuals or, synonymously, particulars. ‘Particularity’, in this context, can be understood in a ‘locational sense’ to mean simply that each event occurs at a time and place and in a literal sense, is, therefore unrepeatable, or it can be understood to distinguish events themselves from the conceptualisations or property descriptions that are imposed upon them. Davidson embraces both senses of ‘particularity’.

The locational sense of particularity is fairly intuitive and easily demonstrated. Davidson (1970a) writes that, “Events are taken to be unrepeatable, dated individuals such as the particular eruption of a volcano, the (first) birth or death of a person, the playing of the 1968 World Series, or the historic utterance of the words, ‘You may fire when ready, Gridley’” (pp.209-210). Of importance to Davidson is that events be understood as ephemeral, that is, short lived or transitory. They have certain durations, after which they cease to be.

Take for example the birth of Jane Austen that took place on the 16th of December 1775 in Hampshire, England. This is a clear instance of an event that is dated, placed, and unrepeatable. Were Jane Austen to be born again, perhaps as a poet, then this would be a second distinct particular event (that is, not just the same event repeated, and not the same event continued), with a different temporal and spatial location from the original birth. Apparent recurrences like the second birth of Jane Austen have led Davidson to emphasise that to claim that an event has recurred, to say that ‘the same thing has happened again’, is usually no more than a careless way of saying that ‘something similar’ has occurred. In other words, “Recurrence may be no more than similar, but distinct, events following one after another” (1970b, p.184).

Similarly, Davidson (1969) urges that it can make no literal sense to say that two events are identical, “… when is one event identical with another? It seems only one answer is
possible: no two events are identical, no event is ever identical with another” (p.163). On the other hand, an event is always identical with itself, by which Davidson means that no matter how an event is described, or redescribed, it remains the same event. Thus, “The death of Scott = the death of the author of Waverley” (1970a, p.210).

The locational sense of event particularity thus focuses upon when and where events occur. Recall that in attempting to give individuating conditions for events, Davidson came to the conclusion that ‘events are identical if they occupy the same places at the same times’. Given that the identification of an event is determined by its spatio-temporal location, it follows that it can make no literal sense to say that an event has been repeated, either elsewhere or elsewhen. Events with distinct spatio-temporal locations are therefore distinct from each other, while apparently distinct events that occupy the same spatio-temporal location are in fact one.

The second sense of event particularity embraced by Davidson is of the utmost importance for resolving the paradox of Anomalous Monism. It involves the distinction between events as they exist in extension and events as they are treated in intension. This distinction itself rests upon a further demarcation proposed by Davidson, the separation of ontological and conceptual categories. As demonstrated above, Davidson insists that events form a fundamental ontological category. By contrast, events types are deemed to form conceptual categories that are intimately concerned with human conceptual and linguistic practices. Thus, ontologically speaking, events exist in extension, that is, in and of themselves, independently of human concerns. They are individuals in the sense that, independently of being conceptualised in one way or another, they do not belong to universal classes or types, such as ‘the mental’ or ‘the physical’. Nevertheless, despite their ontological particularity, Davidson views events as being amenable to conceptualisation and classification into universal types. Whether or not an event is of a certain type, Davidson maintains, is necessarily tied up with its intension, by which he means how that event is ‘picked out’ or described by means of property attributions. Importantly, Davidson (1993a, fn.3) does not distinguish between properties, concepts and predicates. Accordingly, he views the treatment of events in intension to be an inescapably conceptual descriptive practice.
The distinction Davidson draws is thus one between particular events, that is, ontological items that exist and relate independently of human concerns, or, in other words, events that exist and relate in extension, and event types, that is, conceptual items that utilise the property-descriptions or intensionality afforded by interest-sensitive humans, such descriptions providing the material for the formulation of explanations and laws. As seen above, “Causality and identity are relations between individual events no matter how described. But laws are linguistic; and so events can instantiate laws, and hence be explained or predicted in the light of laws, only as those events are described in one or another way” (1970a, p.215).

The extension-intension and the underlying ontological-conceptual distinctions proposed by Davidson can be difficult to grasp and for good reason; it is not at all clear that such distinctions can be maintained. Nevertheless, before contemplating the potential problems that arise from proposing discrepant ontological statuses for events and their property-descriptions, a task that will be addressed in Sections 3.4 and 3.5 below, it is necessary to consider how such a distinction is utilised by Davidson in order to resolve his paradox. The conceptual character of mental property-descriptions and the means by which such descriptions become connected to events will be considered firstly in Section 2.2, and then again in Section 2.5.2.3.

2.2 Constitutive description
As part of his reconciliation of the Principles of Anomalous Monism, Davidson makes what seems a most extraordinary claim, that “… events are mental only as described” (1970a, p.215). In order to understand this claim, and to avoid concluding that it is no more than a tautology (of the kind, ‘we can say that something is X if we can describe it as X’), it is necessary to consider it in the light of the ontological-conceptual distinction that is behind the third person interpretative approach to psychology that Davidson advocates.

As has already been indicated, Davidson’s theory assumes that events exist in extension. Nevertheless, it is also assumed that events are amenable to being described in a variety of ways. It is here that the ontological-conceptual distinction comes into play. Speaking ontologically, Davidson argues that events are particulars. Speaking conceptually, he argues that one and the same event can be described according to one
or more systems of concepts comprising a theory, and that this latter act is what prepares events for admittance into types, “Mental or psychological events are such only under a manner of description, for these very events surely are at the same time neurophysiological, and ultimately physical, events, though recognizable and identifiable within these realms only when given neurophysiological or physical descriptions” (1982b, p.299). As Stephen Noren (1979) sums it up, “… Davidsonian events, events as dated particulars, are deemed mental or physical merely by virtue of how they are described” (p.64). Accordingly, Davidson declares that, “… in my view the mental is not an ontological but a conceptual category… To say of an event, for example an intentional action, that it is mental, is simply to say that we can describe it in a certain vocabulary – and the mark of that vocabulary is semantic intentionality” (1987b, p.46). That is to say, in so far as they are events, perceivings, desirings, actions and so forth are simply located happenings. Only when they are conceptualised according to a psychological theory and described in the appropriate vocabulary do they then become *types* of events, indeed, mental events.

The idea that being an event of a certain type requires only being described in the appropriate vocabulary should not be taken to suggest that descriptions are made arbitrarily, and thus that all events are amenable to all types of description. Events are mental if and only if at least one mental verb is *essential* to the event’s description. Likewise, “Physical events are those picked out by descriptions or open sentences that contain only the physical vocabulary essentially” (1970a, p.211). So how is it to be determined when a vocabulary is essential to the description of an event?

### 2.2.1 Theories and their empirical interpretations

The key, it seems, to conceptualising and describing events in one way or another requires the use of theories. Davidson’s understanding of theories is that they are

---

3 Some authors have argued that while Davidson certainly believes that events are mental only as described, he does *not* claim that events are physical only as described. Skillen (1984), for example, remarks that “Davidson does say that things in themselves are physical; I’ve heard him say that” (p.523), and Antony (1989, 1994) draws attention to the distinction she perceives between Davidson’s empirical approach to physical theories and his conceptual approach to psychological theories. Despite these arguments, it is clear to me that Davidson does intend events to be understood as belonging to one type or another, *only as described*. This is not to claim that description in either the mental or the physical vocabulary is entirely non-empirical (as will be explained shortly, description in one vocabulary or another is deemed to be a product of both theory and observation), rather it is to recognise that Davidson maintains that there exists an ontological-conceptual distinction between events and properties, whether the latter be mental or physical.
composed of *a priori* axioms that guide their usage and determine their range of application. Theories differ from their empirical interpretations in that they are abstract systems that correspond to the interests of their users rather than to the ways of the world. It is only subsequently to the development of a theory that it can be determined whether or not the theory can be applied usefully to the world, “Theories describe abstract structures; their empirical interpretations ask whether these structures can be discovered in the real world” (1995b, p.7).

The divide between abstract theory and worldly application was first encountered in Section 1.2.1 where it was shown that, according to Davidson, the law of transitivity is an axiom of length measurement, and that its place within physical science derives not from empirical observations, but from *a priori* conceptual pressures. The fact that the law of transitivity is taken by Davidson to be partly constitutive of the concept of length was evidenced by his refusal to accept that observation of an intransitive triad of lengths would reveal the law of transitivity to be false. This is what is meant by saying that theories, such as physical theories of length measurement, describe abstract structures. They are deemed to be complete within themselves conceptually, and only possibly having a range of application in the real world. It is prior to attempts to apply them to the world that theories are developed and their axioms asserted.

The conceptual *a priori* nature of theories, as understood by Davidson, carries over to their applications in the world. Theories, he suggests, impose their structures upon the objects and events to which they apply. One of the clearest examples of the imposition of theory upon the world is exemplified by Davidson’s views concerning the role of numbers in measurement. He writes with reference to the weights of objects, “… no one supposes the numbers are in any sense intrinsic to the objects that have weight, or are somehow “part” of them. What are basic are certain relations among objects; we conveniently keep track of these relations by assigning numbers to the objects, and remembering how the relations among the objects are reflected in the numbers” (1989b, p.10). The basic claims within this quotation are that what is real are the relations among the weights of objects (such as being heavier than, equal in weight to, and so forth), and that it is for our more convenient understanding of these relations that we assign numbers to the objects. That weighted objects stand in certain relations to each other is unobjectionable. What is of importance within Davidson’s assertion is the claim
that numbers are assigned to objects in order to serve the interests of people; that their role is one of helping us to ‘keep track’ of the relations between the objects; consequently that numbers are convenient fictions by which people keep abreast of certain types of relations between objects.

The idea that numbers are assigned to objects according to predetermined rules, while popular in psychology, is antithetical to a realist account of the character of numbers and their role in measurement. According to Joel Michell (1999), measurement is “the discovery or estimation of the ratio of a magnitude of a quantity to a unit of the same quantity” (p.222). The ratio that obtains between the first quantity and the second is the number. In other words, numbers are certain kinds of relations between things. It thus follows from a realist account that numbers are discovered by comparing the quantitative attributes of objects, and that both the objects and the ratios of their quantities, that is, the numbers, exist independently of human interests. Consequently, there is no stage at which numbers are assigned either to the objects or to the relations between the objects. Indeed, such assignment is both impossible and unnecessary, because the numbers already exist.

The realist understanding of numbers clearly differs from Davidson’s. Davidson’s view that theories are abstracted from the world, and hence prior to their empirical applications, precludes the possibility that numbers can be discovered in the world because they are assumed to belong not to objects, nor to the relations between objects, but to theories. The role of numbers, it follows, is to serve human convenience, and the place of numbers is within abstract physical theories. Nevertheless, by imposing numbers upon worldly objects, Davidson believes that one can more readily conceptualise the relations between those objects.

Despite what seem to be the obvious pragmatic advantages of assigning numbers to ‘keep track’ of objects’ relations, Davidson concedes that the range of application of numerical assignments relies upon factors beyond human convenience. If they are to be useful, numbers cannot simply be assigned to objects according to whims, rather, the assignment of numbers is constrained in that the objects to be numbered must meet additional criteria before the assignment can take place. One of these, we have seen, is that the objects’ quantitative attributes, such as their lengths and weights, must be
transitive. In the absence of transitivity, assignments of numbers cannot be made. It is thus necessary that a theory be wholly applicable to its proposed objects and events, or not at all. It is this that Davidson is suggesting when he writes that “… we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort…” (1970a, p.221). It is only through satisfying each of the axioms within a theory that an object is amenable to description according to that theory, that is, by the vocabulary appropriate to that theory. Consequently, because an object or event is physical only if described in the physical vocabulary, mental only if described in the mental vocabulary, and so on, the inclusion of an object or event within one kind or another depends on its capacity to conform to the axioms of the theory that attends the vocabulary. In the absence of this conformity, the object or event in question is denied membership in the physical, mental, or other system of description of objects and events.

2.2.2 Mental attributions

Davidson’s views concerning the a priori axiomatic nature of physical theory and the role of numbers therein often serves him as an analogy for the a priori conceptual nature of psychological theory, and the role of an interpreter’s sentences therein.4 In psychology, just as in physics, Davidson believes that we have theories describing abstract structures, and empirical interpretations (or impositions) of those theories in or upon the real world. Behind Davidson’s Anomalous Monism is an abstract theory that utilises the concepts of belief, desire, action and linguistic meaning. This is referred to by Davidson as “… a unified theory of speech and action, or the Unified Theory for short” (1995b, p.8). The empirical interpretation or practical imposition of the Unified Theory upon the world is the practice of radical interpretation.

Conspicuous throughout Davidson’s work is what appears, at first, to be an epistemological focus – the attribution of propositional attitudes to people based upon an interpretation of their behaviour, in particular of their speech. The claim that, “It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual” (1970a, p.222) has been examined already to some extent from the point of view of rational constraints on

---

4 For an argument suggesting that there exists a disanalogy between numerical measurement and propositional attitude attribution, see Davies, 1995, p.53.
the web of propositional attitudes. There is, however, more to this claim than the simple assertion that mental phenomena are responsive to the background in which they are situated. The additional feature concerns the act of *attributing* mental phenomena. What might at first appear to be an epistemological statement about *discovering* what beliefs and so on another person has, can also be read as referring to the act of *assigning* to the person those attributes, just as numbers are assigned to objects. If this reading is correct, then we can perhaps make sense of the idea that ‘events are mental only as described’.

Recall that, according to Davidson, events are particulars and ‘the mental’ is not an ontological but a conceptual category. With this grounding, it soon becomes clear why Davidson frequently employs measurement analogies to help characterise the act of attributing mental contents to people, as well as to suggest what the nature of such attributions might be. The events to which Davidson attributes mental descriptions, if particulars, do not in and of themselves, independently of human concerns, actually have these characteristics. In this they resemble the objects to which numbers are assigned to indicate weight relations. In neither case, according to Davidson, should one be misled into thinking that the assignment of numbers or attitudes is more than a way of talking, as seen by the following, “In thinking and talking of the weights of physical objects we do not need to suppose there are such things as weights for objects to have. Similarly in thinking and talking about the beliefs of people we needn’t suppose there are such entities as beliefs” (1989b, p.11).

If in describing events we are not indicating the characteristics that those events possess, then fairly may it be asked what the point of describing events is. The point, it seems, of making mental attributions is very similar to the point of making numerical assignments. It is to better allow us to ‘keep track’ of the events by imposing upon them an abstract structure that is coherent within itself and readily comprehensible. The analogy between numerical assignments and mental attributions is characterised by Davidson (1989b) as follows, “Just as in measuring weight we need a collection of entities that have a structure in which we can reflect the relations between weighty objects, so in attributing states of belief (and other propositional attitudes) we need a collection of entities related in ways that will allow us to keep track of the relevant properties of the various psychological states” (p.11).
The structured entities referred to in the context of weight are, of course, numbers. What, then, are the entities that apply to psychological states? What do we use in order to keep track of the propositional attitudes of others? According to Davidson (1995b), “Everyone who has a language has available such a set of entities, namely the (infinite) set of his or her own sentences” (p.11). Like the numbers used to assign weights to objects, sentences are used to keep track of the propositional attitudes of others. Of course, it is not just any sentence that Davidson believes can be used to keep track of propositional attitudes, rather, it is in particular those sentences, or utterances, that interpreters use from the mental vocabulary to attribute propositional attitudes to others. He states, “So I will assume we have settled on utterances, the very utterances that are produced in attributing attitudes, as the objects that serve to individuate and identify the various states of mind” (1989b, p.14). That Davidson has settled on interpreters’ utterances to identify and individuate others’ states of mind should not be mistaken, however, for the notion that those minds are in ‘psychic touch’ with the objects (really utterances) that are used to characterise them, or, indeed, with any objects at all. Regarding this he writes that, “My point here isn’t that belief sentences relate believers to sentences, but that this familiar proposal assumes that the objects used to identify a belief may not be within the ken of the believer. Once we grant this possibility, we are free to divorce the semantic need for content-specifying objects from the idea that there must be any objects at all with which someone who has an attitude is in psychic touch” (1989b, p.9).

The attribution of propositional attitudes to people is thus a conceptual operation which aims at making the person to whom the attitudes are attributed more readily understood in the eyes of an observer or interpreter. Through attributing beliefs, desires and meanings to persons by assigning sentences belonging to the mental vocabulary to them, an interpreter treats particular events as if they were events of certain sorts, and in so doing admits those events to the mental class. It is in this literal sense that ‘events are mental only as described’.

Of course, as was mentioned previously, inclusion of events in one class or another is not arbitrary to the extent that any event can be described in any way. Membership in a system of description requires that a comprehensive theory of the object or event being
described holds. The applicability of mental descriptions, Davidson maintains, is determined by the suitability of the Unified Theory to the candidates for description.

2.2.3 The Unified Theory and radical interpretation
The purpose of the Unified Theory is to provide a method for revealing rational connections between an agent’s speech and action on the one hand, and his or her desires and beliefs on the other, from a starting position of complete ignorance of each of these. Underlying the theory is the *a priori* assumption that, ideally, attitudes and actions interrelate in coherent patterns such that adequate knowledge allows for inferences to be made between them, “There are conceptual ties between the attitudes and behavior which are sufficient, given enough information about actual and potential behavior, to allow correct inferences to the attitudes” (1982a, p.322).

The constitutive rationality of the Unified Theory is comprised of two basic elements that were introduced in Section 1.1, holism and normativity. It was seen then that the contents of desires and beliefs, and I may now add the meanings of actions including utterances, are determined in part by the logical relations in which the elements of the web of attitudes and actions stand. The reason for this, though perhaps obscure at that time, derives from the view that such contents are the artefacts of the process of interpreting persons according to a psychological theory, in this case the Unified Theory, in the mental vocabulary. The ‘discovery’ of psychological rationality, Davidson (1990a) writes, “… is an artefact of interpretation, of course, and not an empirical finding” (p.25). Beliefs, desires, actions and meanings are not, it follows, discovered in people, rather they are imposed upon them from the point of view of an interpreter. The interpreter brings to the process some preconceptions of what attitudes and actions are, and how they interact, and then devises an interpretation of the agent’s actions and attitudes so as to achieve a ‘best fit’, a picture of the agent that renders him or her intelligible, according to the interpreter’s notions of what it is to be intelligible. Consequently, the constitutive rationality of propositional attitudes and actions derives not from nature, but from *a priori* maxims contained within abstract psychological theory. It is this feature of psychology that Davidson highlights by saying that, “The limit thus placed on the social sciences is set not by nature, but by us when we decide to view men as rational agents with goals and purposes, and as subject to moral evaluation” (1974, p.85).
The condition that psychological interpretation proceed so as best to reveal the agent being interpreted as holding coherent beliefs, appropriate desires, and subsequently performing actions which are rational in the light of those beliefs and desires, is deemed by Davidson to be an indicator of the creatures to whom psychological interpretation is applicable. He writes that, “Just as the satisfaction of the conditions for measuring length or mass may be viewed as constitutive of the range of application of the sciences that employ those measures, so the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of applications of such concepts as those of belief, desire, intention and action” (1974, p.84). It is therefore only those who are amenable to being described as rational beings that are candidates for being interpreted and thereby attributed beliefs, desires and meanings in accordance with the axioms of the Unified Theory. As has been indicated previously, candidature for being considered rational by Davidson relies upon the possession of a concept of truth, which in its turn is considered to be reliant upon a well developed language ability. The range of application of psychological concepts such as beliefs and so forth is therefore limited, in Davidson’s view, to human beings.

Having demonstrated that underlying the Unified Theory is the assumption of psychological rationality, and that the intended range of application of the theory is to language using humans, it is now necessary to consider the specifics of how, by aiming for a picture of rational coherence, the Unified Theory provides a method for simultaneously interpreting an agent’s utterances, desires and beliefs.

The Unified Theory is derived from Decision Theory as proposed by Frank Ramsey (1926). Ramsey proposes that beliefs have what he calls a ‘subjective probability’, by which he means a numerical degree ranging between 0 and 1 that corresponds with the extent to which the proposition considered is believed. He writes that, “… full belief is denoted by 1, full belief in the contradictory by 0, and equal belief in the two by 1/2” (p.78). In addition, Ramsey suggests that, “… the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it” (p.71). This latter feature provides the key to discovering the degree to which a subject believes in the truth of a proposition, because it gives rise to that subject’s actions as evidenced, for example, in the bets that he or she is willing to undertake with reference to that proposition’s truth. For example, should a person have equal belief in the
propositions ‘this coin will come down heads’ and ‘this coin will come down non-heads’, then he or she will be indifferent to whether winning the bet is attached to the first outcome or the second, that is, between the options ‘winning if heads turn up, losing if heads do not turn up’ and ‘losing if heads turn up, winning if heads do not turn up’.

The conclusion that a person’s accepting a bet or embarking on some other course of action reveals his or her degree of belief in a proposition that is relevant to that action assumes, of course, that he or she will act so as to satisfy his or her desires. It is not only a subject’s degree of belief, therefore, but also his or her preferences or values, that determines the choices that he or she makes, and the actions that he or she engages in. Evident in the actions, including choices, that a person engages in, therefore, should be the degree of belief that he or she accords various propositions that can be acted upon, and also the pattern of preferences for the potential outcomes of so acting that are foreseen. When the outcomes are obviously comparable with each other, as occurs for example with monetary bets involving differing amounts, then the task of estimating the numerical values of preferences and beliefs is somewhat straightforward, at least in comparison to when the outcomes are less obviously comparable with each other, as occurs, to use Ramsey’s example, between such outcomes as opportunities for an hour’s swimming and an hour’s reading. Nonetheless, Davidson credits Ramsey’s theory with having shown that, “… if the pattern of an individual’s preferences or choices among an unlimited set of alternatives meets certain conditions, then that individual can be taken to be acting so as to maximize expected utility, that is, he acts as if he assigns values to the outcomes on an interval scale, judges the plausibility of the truth of propositions on a ratio scale, and chooses the alternative with the highest computed expected yield” (1974, p.82).

While Decision Theory provides much of what Davidson needs for a model of rational connections between belief, desire and action, it is lacking in one important respect, namely, it does not involve a consideration of linguistic meaning. In proposing the Unified Theory, Davidson seeks to improve upon Decision Theory by adding to it a theory of meaning, with the resultant Unified Theory being, as its name suggests, a unified theory of speech and action. The following paraphrase of the methodology of the Unified Theory is taken from Davidson (1995b).
The Unified Theory suggests that by ascertaining an agent’s preferences through the simple choices that he or she makes, a rational interpretation of his or her words, desires and beliefs can be developed. In accordance with Decision Theory, numbers are assigned to beliefs and desires in such a way as to preserve the qualitative constraints that are imposed by the theory. These constraints are such that the overall picture that emerges blends the attitudes and actions of the agent so as to render them internally consistent. Through observation of the relations between the agent’s preferences for one sentence over another, inferences can be made concerning how much belief the agent has in his or her sentences, the extent to which the agent would like those sentences to be true, and the truth conditions or meanings of those sentences. Subsequently to these inferences being made, predictions of actions can be made from the patterns of desires and beliefs that are represented by the numbers. The intended outcome of applying the theory to an individual is an interpretation of his or her beliefs, values and words based simply upon the observed patterns of preferences that he or she has for certain sentences.

According to Davidson (1995b), putting the Unified Theory into practice can be viewed either ‘officially’, that is, according to how it would be operationalised ideally, or ‘unofficially’, that is, how it is actually operationalised. The ‘official story’ is the process of radical interpretation, which involves the following steps:

The fact that a rational agent cannot prefer both a sentence and its negation to a tautology, or a tautology to both a sentence and its negation (by analogy with the coin example above) allows, Davidson claims, an interpreter to identify the sentential connectives within an agent’s sentences (these including negation, conjunction and biconditionals) whilst remaining ignorant of the meanings of those sentences. From this minimal amount of knowledge the interpreter determines how likely the agent thinks his or her sentences are to be true as well as his or her preference values for the truth of those sentences, both in numerical terms, and then, working with the subjective probabilities, proceeds to interpret the agent’s sentences, that is, to assign them content.

At this stage the numbers that have thus far represented the attitudes, actions and utterances are replaced by the interpreter’s sentences which function in much the same way as the numbers did, modelling the attitudes and thereby keeping track of the
thoughts, words and actions of the agent (see Section 2.2.2). Content-giving interpretation begins with what Quine has called ‘observation sentences’, which are those that refer to events and objects that can be perceived simultaneously by both the agent and the interpreter (this will be returned to below under the heading of ‘triangulation’). The final step involves the interpretation of sentences other than observation sentences, through registering the entailment and conditional probability relations between these and the observation sentences.

Both the theory and its official interpretation are, Davidson (1995b) claims, remote from actual practice. This divide between theory and practice he nevertheless deems to be unproblematic because, “The point of the theory is not to describe how we actually interpret, but to speculate on what it is about thought and language that makes them interpretable. If we can tell a story like the official story about how it is possible, we can conclude that the constraints the theory places on the attitudes may articulate some of their philosophically significant features” (p.10).

Interpretation and the attribution of beliefs, desires and meanings in accordance with the constraints of the Unified Theory ‘reveal’, or, more accurately, prescribe, the features that psychological attitudes will be treated as having. Two of these, holism and normativity, have been revealed as central to the coherent assignment of contents to an individual. A third that has been met with already, externalism, is likewise an essential tenet of the Unified Theory, both for its role in the assignment of content to beliefs and utterances, and, just as importantly, as a means of justifying the practice of assigning such contents.

The externalist element of the Unified Theory is essential for radical interpretation because it underlies ‘observation sentences’. While the truth conditions of individual sentences are in general deemed to be reliant upon the logical relations between these and other sentences that the agent utters, observation sentences are constrained additionally by their causal histories with the external world. Just as beliefs are ‘external’ in the sense that they are partially defined in terms of the objects and events that cause them, so too observation sentences are ‘external’ in that their interpretation requires the interpreter’s taking into account the objects and events to which they refer. The attribution of meaning to another’s sentences is a process that Davidson (1973b)
deems to be domestic as well as foreign, which is to say, that radical interpretation is equally necessary whether the person interpreted is apparently speaking the same language as the interpreter or not. The unsupported assumption that others use words in precisely the same way that we, as interpreters, do, is, according to Davidson, without justification. In order to ascertain that a word or sentence issuing from you does in fact have the same meaning as when it issues from me requires what Davidson terms the practice of ‘triangulation’.

In proposing the method of triangulation, Davidson is aiming to illuminate how it is that one can come to know what another thinks and means, from a third person perspective. He writes that, “Externalism makes clear how one person can come to know what someone else thinks, at least at the ground level, for by discovering what normally causes someone else’s beliefs, an interpreter has made an essential step toward determining the content of those beliefs. It is not easy to conceive how else it would be possible to discover what someone else thinks… the interpreter must know, or correctly surmise, the events and situations that cause a verbal or other reaction in another person in order to fathom her thoughts…” (1989a, p.195). 5

Davidson is suggesting that the process of interpreting another requires the ability to perceive what it is that simultaneously causes the other person’s thoughts, and provides the referent for his or her utterances. He proposes that as a condition of thought there must exist three interacting classes of objects and events: a perceivable external situation, a response to that external situation by the interpreter, and a response to that situation by the person interpreted. Taking as an example a child who responds to a table by uttering ‘table’, Davidson suggests that an interpreter determines the contents of that child’s belief, and simultaneously the meaning of its utterance, by discovering the stimulus that causes both. He writes that, “It’s a form of triangulation: one line goes from the child in the direction of the table, one line goes from us in the direction of the table, and the third line goes from us to the child. The relevant stimulus is where the lines from child to table and from us to table converge” (1989a, p.198).

5 A similar conclusion is reached in Davidson’s 1990b consideration of Turing’s test, although the focus there is on discovering that, not what, another thinks.
The process of triangulation, Davidson (1989a) emphasises, is not simply a means of identifying the objects and events to which an interpreted creature responds. Indeed, he has argued elsewhere that a creature’s merely responding to or discriminating between features of the world differs from its having a belief, as evidenced by the discriminatory powers of sunflowers and earthworms, neither of which Davidson (1996) considers to have thoughts. Rather, the process of triangulation is viewed by Davidson as a necessary but not sufficient condition of thought. Accordingly, he urges that “… there could not be thoughts in one mind if there were no other thoughtful creatures with which the first mind shared a natural world” (1989a, p.193).

The suggested impossibility of thought in an unobserved mind stems from the problematic nature of determining which objects and events that mind is responding to. In the presence of a triangle, intersecting lines from the two people to the object allow for identification of that object. By contrast, a single line projecting from a mind outwards offers no means of fixing the object that determines the content of a belief or utterance. Under these circumstances, Davidson argues that there would be no more reason to claim that the object was a table than that it was something more proximal to the person, such as stimulation of his or her sensory nerves, and the question of what object caused, and thus gave content to a thought, would be in principle unanswerable, “… until the triangle is completed connecting two creatures, and each creature with common features of the world, there can be no answer to the question whether a creature, in discriminating between stimuli, is discriminating stimuli at the sensory surfaces or somewhere further out, or further in” (1996, p.166). The conclusion that Davidson considers as inevitably arising from the uncertain fixation of objects of thought is that there is in this scenario no reason to maintain that the solitary creature has any thoughts at all. He thereby concludes that “… if someone has thoughts, there must be another sentient being whose innate similarity responses are sufficiently like his own to provide an answer to the question, what is the stimulus to which he is responding?” (1989a, p.199).

Externalism is therefore fundamental to the Unified Theory because it suggests to an interpreter what contents to attribute to the thoughts and utterances of others. Without this content establishing step, the process of radical interpretation would be without grounding, firstly because there would be no reason to believe that others had thoughts,
and secondly because no one internally consistent interpretation would have any basis for being considered better than any other.

The Unified Theory and its official interpretation paint a picture of psychology whereby practitioners, under which label falls any person who attempts to understand another, interpret their fellows in such a way as to reveal them as thinking and acting in perfect rationality. Further, this interpretation is accomplished using only the evidence gained from observing the interpretee’s preferences and observation sentences, whilst observing the norms for interpretation that are dictated by the prescribed holism and normativity within the theory. Nevertheless, Davidson acknowledges that real life interpretations, whilst guided by the theory, differ both in their methods and the degree to which they can conjure a rational picture of an agent’s thoughts and actions. He writes that, “Unofficially, one can admit that as living, working interpreters, we never have enough of the sort of evidence needed to follow the official route, and we always have a great deal of other sorts of evidence. We make endless assumptions about the people we meet, about what they want, what they are apt to mean by what they say, what they believe about the environment we share with them, and why they act as they do. Our skills as interpreters come into play mainly when one or another of these assumptions turns out to be false, and by then we have much more than the poverty-stricken evidence the Unified Theory depends on” (1995b, p.10).

The unofficial real life process of interpreting others differs from the abstract official process of radical interpretation in a number of ways. Of importance are the dynamic and charitable characters of the real life process, as opposed to the static and exact characters of the official process.

The Unified Theory, Davidson (1995b) tells us, “… is static; it says nothing about the forms of rationality that deal with the incorporation of new information into a going system of thought” (p.15). Consequently, the official mode of interpretation is unable to cope with changes of beliefs and desires across time. By contrast, real life interpretation, if it is to be successful, must be dynamic, that is, amenable to alteration as new considerations arise. In other words, in real life interpretation there is always the possibility of faulty interpretation resulting from the inadequacy of present-moment evidence to capture the dynamic process of thought and action across time.
The second important difference between the official and unofficial processes of interpretation is that while the former has no means of coping with irrationality or error, the latter must incorporate anomalies as best it can. Highlighting this discrepancy, Davidson writes that the formal theory “… says nothing at all about inconsistencies. It not only postulates perfect logic and a consistent and rational pattern of beliefs and desires, but it assumes rationality in the treatment of what we take to be evidence. Inconsistencies and failures of reasoning power must be accommodated by injecting large doses of what has been called charity in the fitting of the theory to actual agents” (1995b, p.14).

A perfect interpretation of the thoughts, meanings and actions of another would have no need to incorporate any false beliefs or irrational actions because the contents and meanings attributed to these would accord perfectly with each other. Nevertheless, real life interpretation proceeds less smoothly than could be desired. As interpreters, we encounter elements within others’ psychological makeups that simply cannot be reconciled with each other, or with the objects and events which supply the content of those others’ beliefs and utterances. At these times we must employ the ‘Principle of Charity’. The Principle of Charity is a methodological guideline that instructs us to interpret others so as to render them as rational, correct, and hence intelligible as possible. This raises the possibility that a choice will have to be made between interpreting the sayings of others according to language as we use it, or, alternatively, interpreting others as using sentences that sound the same as our sentences whilst quite different meanings are intended, should the situation demand it. The decision, Davidson (1968) admits, is a difficult one. He writes that, “Hesitation over whether to translate a saying of another by one or another of various non-synonymous sentences of mine does not necessarily reflect a lack of information: it is just that beyond a point there is no deciding, even in principle, between the view that the Other has used words as we do but has more or less weird beliefs, and the view that we have translated him wrong. Torn between the need to make sense of a speaker’s words and the need to make sense of the pattern of his beliefs, the best we can do is choose a theory of translation that maximises agreement” (p.101).

---

In maximising agreement between ourselves and those we interpret we attribute to them
the most rational pattern of meanings and beliefs that we can. Part of this interpretation
may require the attribution of intelligible error or occasional irrationality. The Principle
of Charity directs us, in these instances, to render the other as rational and as correct
about the world as is possible, as seen in the following, “The best we can do is cope
with error holistically, that is, we interpret so as to make an agent as intelligible as
possible, given his actions, his utterances and his place in the world. About some things
we will find him wrong, as the necessary cost of finding him elsewhere right. As a
rough approximation, finding him right means identifying the causes with the objects of
his beliefs, giving special weight to the simplest cases, and countenancing error where it
can be best explained” (1983, p.318).\footnote{As an additional means of coping with irrationality, Davidson (1982b, 1985c) accepts that the mind may be conceived to be partitioned as suggested by Freud. Such a solution, Davidson recognises, is at odds with the motivation for using the Principle of Charity, because “… the point of partitioning was to allow inconsistent or conflicting beliefs and desires and feelings to exist in the same mind, while the basic methodology of all interpretation tells us that inconsistency breeds unintelligibility” (1982b, p.303). Further discussion of how best to understand irrationality is contained in Davidson (1970c).}

The Unified Theory, thought of only in the abstract, is a system of axioms that
prescribes the rational coherence of psychological attitudes and actions both with each
other and with the surrounding environment. When imposed upon people, the theory
allows for holistic interpretation of what people say, do and believe, all in terms of what
it is reasonable for them to say, do and believe. Importantly, to the extent that
candidates conform to the axioms of the theory, they are amenable to description in the
mental vocabulary, and hence to the attribution of psychological categories. When the
axioms of the theory are violated too greatly, when no rational pattern of thought and
action can possibly be attributed to the candidate, then description in the mental
vocabulary is denied. Thus, whether a creature has mental attributes or not depends
wholly on whether that creature is amenable to depiction through the Unified Theory,
and, subsequently, whether that creature can be described in the mental vocabulary.
Such is Davidson’s view that ‘events are mental only as described’.

2.3 Causal relations and causal explanations

Through postulating the extension-intension distinction, whereby ontologically
speaking events are particular but conceptually speaking events can be described in
terms of properties and hence types, Davidson is likewise able to draw a distinction
between causal relations and causal explanations, including laws. Like identity, causality ‘deals with events in extension’, which is to say that causal relations occur between individual events, indifferently to how they are described. Causal explanations, including laws, on the other hand, are thought by Davidson to be linguistic entities that are sensitive to the interests of the people postulating them. This is because explanations are formulated in terms not of individual events, but of types of events. It follows from the view that the ‘properties’ of events that gain them membership in a type are in reality artefacts of descriptive practices, that laws and other explanations, which are formulated in terms of event types, are likewise about events as described in one way or another, not about the events themselves. Consequently, to understand Davidson’s position it is necessary to observe the distinction between causal relations and the singular causal statements that refer to them on the one hand, and causal explanations and laws on the other. These will be considered in turn.

2.3.1 Causal relations and singular causal statements

Beyond claiming that causal relations obtain between individual events no matter how those events are described, Davidson provides little analysis of causation. What can be gleaned, however, from the claim that causation is a relation between individual events is that one event’s causing of another occurs indifferently to human interests. Specifically, how events are described by humans has no impact upon their causal efficacies.

Because he takes events to be concrete individuals, Davidson believes that it is possible to give them what he calls ‘definite descriptions’, such as names. Just as each individual person has a name that is unique to him or her, and by which he or she can be specifically referred to, Davidson believes that each event can be named by a ‘singular term’. As Evnine (1991) points out, few events have names in the way that people do, so the singular terms that are used to refer to events more typically take forms such as ‘the sinking of the Titanic’ and ‘the death of Shakespeare’. The most obviously important feature of a singular term is that it is definite. This is what distinguishes ‘the explosion’ from ‘an explosion’; the first, but not the second, uniquely names an event. In addition, individual events can be referred to by more than one singular term, such terms being co-extensive. For example, ‘the Sydney Olympic Games’ and ‘the Games of the XXVII Olympiad’ are co-extensive, they each refer to the same event.
Singular causal statements make use of singular terms in that they name two events, a cause and an effect, and additionally claim that the first event caused the second. For example, ‘the squashing of the fly caused the discolouring of the wall’ names two events (the squashing and the discolouring), and asserts that the former caused the latter. Singular causal statements are true when the first event referred to did indeed cause the second, and false when it did not. The truth or falsity of singular causal statements is determined, in other words, by the events themselves. Importantly, the truth of a singular causal statement cannot be altered by the substitution of co-extensive or co-referring singular terms. Thus, if ‘the squashing of the fly’ is co-extensive with ‘the killing of the fly’, then substitution of the second singular term into the above singular causal statement will not alter its truth value. This stability of the truth value occurs because “Singular causal statements are extensional: their truth value is invariant under the substitution of one name or description of an event for another name or description of the same event” (1995a, p.265).

2.3.2 Causal explanations

Singular causal statements, it has been shown, serve no greater purpose than to state that one event caused another. What is generally of more interest to observers, however, is the question of why an effect has been brought about. It is the role of explanation to satisfy one’s curiosity concerning the ‘why’ question.

Underlying the ‘that-why’ distinction between singular causal statements and causal explanations is the difference that, while the former are extensional, the latter are intensional (i.e., sensitive to how the events being explained are described). Thus, while a singular causal statement can do no more than impersonally name the events involved in a causal relation and state that one caused the other, a causal explanation can offer descriptions of the events that satisfies one as to how they fit into a familiar picture. Causal explanations proceed, it follows, by describing causally related events in one vocabulary or another and then appealing to theories concerning events so described.

Importantly, Davidson maintains that because explanation is intensional, multiple explanations can be given simultaneously for singular causal relations. Consider the following example, “Suppose Magellan notices that there are rocks ahead, an event that, through the intervening events such as his uttering orders to the helmsman, causes the
ship to alter course. Magellan’s noticing is a mental event, and it is causally efficacious. That event is also a physical event, a change in Magellan’s body, and describable in the vocabulary of physics” (1993a, p.12). The one event, Davidson claims, is amenable to description in both the mental and the physical vocabulary (i.e., as a noticing and as a bodily change), and is therefore suitable for inclusion in both a psychological and a physical system of explanation.8

A psychological explanation of why Magellan acted as he did refers to his noticing the rocks ahead, wanting to avoid a collision, and acting accordingly. Described in these terms, that is, as reasons and actions, what Magellan did fits into the familiar picture of rational human thought and action that is presupposed by theories such as the Unified Theory. Nevertheless, the same events, Davidson maintains, are simultaneously amenable to description in the physical vocabulary. The event that can be described as a ‘noticing’ can also be described as ‘a change in Magellan’s body’. An action such as uttering orders could be described in terms of the motions of the lips and tongue. When described in physical terms such as these, the events formerly described as thoughts and actions can now be incorporated into a physical system of explanation.

Formulated in terms not of objects and events themselves, but of interest sensitive descriptions of those objects and events, Davidson understands explanations to be intensional in the sense that they deal with objects and events as described in one way or another. Due to their reliance on linguistic attributions of mental, physical or other properties, explanations are considered by Davidson to be a way of speaking about individual events in familiar, universal, terms. In so doing, it is at times possible to refine explanations to such an extent that a strict law emerges, however, as will be recalled from Section 1.2, such refinement can occur only when the events referred to are described physically. Described mentally, the same events lend themselves to explanations that need not, and indeed cannot, be suitably refined for inclusion in strict

---

8 The notion that multiple explanations are applicable to singular instances of causation has been argued against repeatedly by Kim (1988, 1989a, 1995), who advocates the ‘principle of explanatory exclusion’, i.e., the principle that, “… there can be at most one complete and independent explanation of a single explanandum” (1995, p.126). Davidson rejects Kim’s criticism by re-emphasising the distinction between extensional causal relations and intensional explanations as follows, “It is only if we confuse causal relations, which hold only between particulars, with causal explanations, which, so far as they are ‘sufficient’ must deal with laws, and so with types of events, that we would be tempted to accept the principle of causal-explanatory exclusion” (1993a, p.16). The crucial distinction between the two positions turns on Davidson’s assumption, which Kim rejects, that properties are conceptual in nature. To understand and criticise Davidson’s theory effectively requires that this assumption is not neglected.
laws and that, furthermore, cannot be connected to physical explanations, “… the mental and physical can be construed as two autonomous nomological systems for describing the same events. On this view, there are no psychophysical laws, no law-like statements ‘across’ such systems” (Noren, 1979, p.67).

Despite this, Davidson (1967b) argues that strict deterministic laws play an essential role in justifying one’s belief in any singular causal statement, “…we have reason to believe the singular statement only in so far as we have reason to believe there is such a law” (p.160). So where, exactly, do strict laws fit in to causal explanations formulated in mental terms? According to Davidson, it is not necessary to know the particular strict law that covers a singular causal statement, rather, it suffices to believe that there exists such a law. This means that, for any singular causal statement such as ‘a caused b’, there exist descriptions of the events a and b that, if known, could be refined such that they could appear as terms in a strict law. Further, given this law, and the statement that either of the events, so described, occurred, it could be deduced that the other event also occurred.

Thus, while it may be assumed that where there is a causal relation, there is a covering strict law, it is an error, Davidson (1963) argues, to assume that “…singular causal statements necessarily indicate, by the concepts they employ, the concepts that will occur in the entailed law” (p.684). A singular causal statement such as ‘the belief that it would rain caused the man’s closing of the windows’ employs the concepts of belief and action stated in the mental vocabulary. If Davidson is correct, then this causal relation will be covered by a strict deterministic law, but, importantly, the strict law will not make use of the mental descriptions that were originally employed by the singular causal statement. Instead, it will be the events’ redescriptions, their physical descriptions, that will indicate the relevant law. Therefore, for any singular causal statement that is amenable to explanation in mental terms, there must exist alternative descriptions of the cause-event and the effect-event that reveal the singular causal relation to be an instance of a strict deterministic law. It is this second set of descriptions that, by showing the singular causal relation to be covered by a strict deterministic law, supplies the justification for belief in the relation being a causal one.
What Davidson (1967b) insists should not be concluded from the above line of reasoning is that whether an event is causally efficacious or not is in some sense dependent upon how it is described and, in particular, whether it is described in such a way that it indicates the relevance of a strict law. He writes that, “… we must distinguish firmly between causes and the features we hit on for describing them, and hence between the question whether a statement says truly that one event caused another and the further question whether the events are characterized in such a way that we can deduce, or otherwise infer, from laws or other causal lore, that the relation was causal” (p.155).

Davidson reminds us here that events themselves are particular, and that the relations they enter into occur indifferently to our descriptions. It is only when we come to describe events for the purposes of explanation and the formulation of laws that descriptions, such as mental and physical descriptions, become relevant. Consequently, it can be of no relevance to an event’s causal efficacy that it is described in such a way that the laws pertaining to it are less than strict and exceptionless. What does occur, however, is that the explanations resulting from such descriptions are ‘low-grade’, having more to do with tendencies and probabilities than with certainty and predictability. Nevertheless, Davidson (1963) argues, most causal explanations are less than perfect and this need not prove problematic, rather, “Ignorance of competent predictive laws does not inhibit valid causal explanation, or few causal explanations could be made” (p.683).

In summary, causal relations obtain between individual events independently of those events’ descriptions. Causal explanations, of which causal laws are a species, are formulated in interest-dependent vocabularies. Whenever a causal relation occurs, there are appropriate (physical) descriptions of the events involved which are amenable to inclusion in a strict deterministic law. There may also be non-physical descriptions of one or more of the events that can be included in less strict but still useful causal explanations. Included in the latter group of explanations are psychological rationalisations.
2.4 Physical-mental event token identity
Magellan’s noticing rocks ahead is, Davidson claims, both a physical and a mental event. Elsewhere, he has claimed that “… mental events are identical with physical events” (1970a, p.209). It should now be much clearer as to what is driving Davidson’s psychophysical monism: individual events are amenable to being described in both mental and physical vocabularies and, therefore, taken one by one, each mental event is also a physical event.

As proof of his monism, Davidson employs an interesting line of reasoning that involves a cross-over between extensional causal relations and intensional causal explanations. The argument turns upon the assumptions that mental events are causally efficacious, that where there is causality there is a law, and that laws must be formulated in the physical vocabulary. It proceeds as follows: If each causal relation involves a law and only events described in physical terms can be included in laws, then all events entering into causal relations must have appropriate physical descriptions. Mental events enter into causal relations, therefore mental events must have physical descriptions, in which case mental events are physical events. Hence mental anomalism and mental efficacy together imply psychophysical monism.

There remains, however, the need for further elucidation of the idea that an individual event can be simultaneously described according to both mental and physical concepts, without there then being a lawful connection between types of mental and physical concepts or descriptions. Needless to say, if an event’s physical description is receptive to refinement into a predictive and explanatory law, and its mental description is lawfully connected to its physical description, then it is a short passage from events having both mental and physical descriptions to there being mental descriptions that appear in predictive and explanatory laws. It has already been seen that Davidson denies there being any lawful connection between an event’s mental description and its physical description. Indeed, he claims that persons in identical physical conditions may well differ psychologically (see Section 1.2.2.1). The argument that Davidson relies upon for establishing the independence of an event’s various kinds of description is borrowed from the concept of ‘supervenience’ that has been employed in ethics. Such is Davidson’s reliance upon supervenience that he admits that Anomalous Monism and the principles from which it is derived may be refuted if they can be shown to be
inconsistent with the supervenience of mental properties upon physical properties (1993a). With this admission in mind, it is important to review the concept of supervenience as it first appeared in ethics, and the way in which it has been adapted for use in psychology.

2.5 Supervenience

The term ‘supervenience’ is derived from the Latin supervenio whose various translations include ‘to come upon or over’, ‘to come after, to follow’, ‘to exceed, go beyond’ and ‘to surpass, exceed, excel’ (White & Riddle, 1880, p.1903). These ‘temporal’ and ‘exceeding’ senses of supervenience have within recent philosophical works given way to a new sense, one suggestive of ‘dependence’. It was this latter sense of supervenience that Davidson encountered upon reading the work of a philosopher of ethics, Richard Hare (Davidson, 1993a).

From morals to mind, ‘supervenience’ has retained its connection with ‘dependence’, and it is hoped by Davidson that this form of dependence will allow mental events, taken individually, to be identified with physical events without a concomitant requirement for lawful connections between types of mental and physical events. In this manner, Davidson hopes to establish an identity theory of mind which encompasses the necessary rationality and anomalism of propositional attitudes, whilst demonstrating the causal efficacy of such attitudes. He writes that,

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect (1970a, p.214).

Clearly supervenience is envisaged by Davidson to involve some sense of dependence, whereby the existence of various combinations of physical characteristics is a necessary condition for the existence of mental characteristics. Given the denial of psychophysical laws, the type of dependence obviously cannot extend to sufficiency as well. The question, then, is whether a property of one sort can be dependent on properties of another in that in order for the former to exist, so too must the latter, without this relation being a lawful one, one that connects types of mental and physical properties.
Before addressing this question, however, it would be well to consider the supervenience relation in the context that Davidson discovered it, that is, in the writing of Hare (1961).

2.5.1 Supervenience in ethics and aesthetics

2.5.1.1 Subvenient and supervenient properties

Hare (1961) is concerned with ethics and the making of value judgements. He takes pains to demonstrate that there is a special feature that characterises value words and which distinguishes this group from those used merely to describe. He writes that, “it is a feature sometimes described by saying that ‘good’ and other such words are the names of ‘supervenient’ or ‘consequential’ properties” (p.80). Consider Hare’s example concerning the supervenient property ‘good’,

Suppose that a picture is hanging upon the wall and we are discussing whether it is a good picture; that is to say, we are debating whether to assent to, or dissent from, the judgement ‘P is a good picture’… First let us notice a very important peculiarity of the word ‘good’ as used in this sentence. Suppose that there is another picture next to P in the gallery (I will call it Q). Suppose that either P is a replica of Q or Q of P, and we do not know which, but we do know that both were painted by the same artist at about the same time. Now there is one thing that we cannot say; we cannot say ‘P is exactly like Q in all respects save this one, that P is a good picture and Q not’. If we were to say this, we should invite the comment, ‘But how can one be good and the other not, if they are exactly alike? There must be some further difference between them to make one good and the other not.’ Unless we at least admit the relevance of the question ‘What makes one good and the other not?’ we are bound to puzzle our hearers; they will think that something has gone wrong with our use of the word ‘good’. Sometimes we cannot specify just what it is that makes one good and the other not; but there always must be something. Suppose that in the attempt to explain our meaning we said: ‘I didn’t say that there was any other difference between them; there is just this one difference, that one is good and the other not. Surely you would understand me if I said that one was signed and the other not, but that there was otherwise no difference? So why shouldn’t I say that one was good and the other not, but that there was otherwise no difference?’ The answer to this protest is that
the word ‘good’ is not like the word ‘signed’; there is a difference in their logic (pp.80-81).

The ‘difference in logic’ between descriptive words such as ‘signed’ and value words such as ‘good’ is accounted for by the fact that while descriptive properties ‘stand alone’ in an object, that is, exist independently of other descriptive or value properties, value properties exist in virtue of descriptive properties. This is what is meant by calling value properties ‘consequential’. An object’s having a value property such as ‘goodness’ is a consequence of its having certain other, descriptive, properties.

To expand upon Hare’s example, imagine that a critic has just asserted that ‘P is a good picture’ and that when questioned as to why she has commended the picture in this way, she replies that the picture has symmetry (S), realistic depth (D), and bright colours (C). Assuming, for the sake of simplicity, that the critic is aware of why she has judged the picture to be good, that is, that there are no other features of the picture that contributed to her judgement, then it follows that the value-property ‘good’ of the picture is determined by, or a consequence of, its descriptive features (S, D and C). In terms of supervenience, the value property (goodness) of the picture is the supervenient property, and the descriptive properties upon which the goodness depends are the subvenient properties.

The above sense of supervenience is known in contemporary philosophy as ‘weak supervenience’, which Jaegwon Kim (1984a) defines as follows, “A weakly supervenes on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A - that is, indiscernibility with respect to B entails indiscernibility with respect to A” (p.58).

The ‘A’ and ‘B’ in Kim’s (1984a) definition refer to the set of supervenient properties and the set of subvenient properties respectively. In the picture example, the set A contains two properties: (i) being good and (ii) being non-good. Set B contains all possible conjunctive combinations (of which there are eight) of the presence and absence of symmetry (S), realistic depth (D), and bright colours (C). The individuals, x and y, correspond to the two pictures, P and Q. The purport of Kim’s definition is that if the pictures share the same combination of factors from set B (e.g., they are
symmetrical but lack realistic depth and bright colours) then, if set $A$ weakly supervenes on set $B$, they will also share the same factor from set $A$ (that is, it must be the case that either both pictures are good or that neither of the pictures is good). This is what is meant by saying that indiscernibility with respect to subvenient properties entails indiscernibility with respect to supervenient properties. This definitional feature of supervenience I shall refer to as the ‘indiscernibility requirement’.

The indiscernibility requirement of supervenience accounts for why Hare’s audience would have been puzzled by the claim that while $P$ and $Q$ are identical, $P$ is a good picture but $Q$ is not. By definition, if goodness is supervenient on symmetry, realistic depth, and bright colours, then it cannot be the case that there exists one picture with these features which is good, while another picture embodies these features, but is not good. When the subvenient descriptive properties agree, so too must the supervenient value property.

The indiscernibility requirement of supervenience is not, it should be noted, bi-directional. That is, indiscernibility of supervenient properties does not entail indiscernibility of subvenient properties. Consider two pictures, $P$ and $T$ that do not resemble each other in the least. Our critic may judge both pictures to be ‘good’, $P$ because it has symmetry, realistic depth, and bright colours, and $T$ because it is large and thickly textured. The indiscernibility requirement is such that if another picture, $V$, were large and thickly textured, then it, too, must be good. However, there may be any number of good pictures which have no descriptive properties in common whatsoever. Therefore, the indiscernibility requirement operates from subvenient properties to supervenient properties, not from supervenient properties to subvenient properties. Given a combination of subvenient properties, the supervenient property is determined. Conversely, given a supervenient property, one or other of the appropriate combinations of subvenient properties has obtained.

2.5.1.2 The standard in supervenience

The present characterisation of supervenience tells only part of the story about the relationship between supervenient and subvenient properties. As a result, the supervenience relation may well appear to be mysterious. What we have seen so far is that value properties exist in virtue of descriptive properties, that is, that they are
dependent upon descriptive properties for their existence. This dependence of value properties upon descriptive properties might have some sort of intuitive appeal – it is always legitimate to ask why an object has been evaluated in a certain way, and this presumes that there is a reason for something’s being good or otherwise. However, in proceeding straight to what are really implications of supervenience, such as that if a combination of subvenient properties determines a supervenient property in one instance, it will do so in a second instance too, we neglect what is really a prior task, that of specifying why it is that certain combinations of subvenient properties are linked to particular supervenient properties while other combinations are not. For instance, why was it the case that upon observing the descriptive characteristics (S, D and C) of picture P, the critic judged the picture to be ‘good’ instead of ‘bad’? In what sense are symmetry, realistic depth and bright colours linked to a picture’s being good rather than bad or even indifferent? In other words, what accounts for this mysterious connection between supervenient and subvenient properties?

One possibility that suggests itself in response to the apparently mysterious nature of the link between subvenient and supervenient properties is that there is some kind of necessary connection between the two. Of this possibility Hare (1961) writes that, “… a natural response to the discovery that ‘good’ behaves as it does, is to suspect that there is a set of characteristics which together entail a thing being good, and to set out to discover what these characteristics are. This is the genesis of that group of ethical theories which Professor Moore called ‘naturalist’…” (pp.81-82).

The necessity involved in a naturalist approach to understanding supervenient properties may be of a definitional sort, whereby it is possible to ‘reduce’ supervenient properties to their subvenient determiners because, in evaluating an object, we mean only to say that it has certain descriptive properties. Alternatively, the necessity may be understood as metaphysical. This is to say that supervenient properties necessarily supervene upon subvenient properties, that this is a ‘necessary truth’ about the world. Such ‘necessity’ might or might not waylay concerns about the mystery of supervenient dependence, but regardless of this there are reasons for viewing naturalist accounts of supervenience with suspicion.

---

Hare (1961) rejects the naturalist approach to supervenience in ethics, arguing that it is misguided. The fundamental flaw with naturalist approaches, Hare argues, is “… that they leave out the prescriptive or commendatory element in value-judgements, by seeking to make them derivable from statements of fact” (p.82). In other words, if by judging something to be ‘good’ we mean only to say that it has a certain set of subvenient properties, then the primary purpose of calling something ‘good’, that of commendation, is lost. No value judgement has been made.

In this way Hare (1961) draws a distinction between terms being **descriptive** and being **evaluative**, an example of the former being ‘This strawberry is large, red, and juicy’, and the latter ‘This is a good strawberry’. The important point to understand is that, “The first sort of remark is often given as a reason for making the second sort of remark; but the first sort does not by itself entail the second sort, nor vice versa” (p.111). In other words, while it may be the case that a strawberry is a good one because it is large, red and juicy, its being good cannot be reduced to its possession of these properties. If such reduction were possible, if judging a strawberry to be good were simply a shorthand way of describing it as large, red and juicy, then, in saying that a strawberry is good, we simply describe it, we do not judge it.

Thus, Hare’s (1961) rejection of a naturalistic interpretation of supervenience is founded primarily on the distinction observed between describing something and evaluating that same thing. At least two problems arise from the attempt to reduce evaluative properties to descriptive ones. Firstly, if by calling something ‘good’ we mean simply that it has a certain set of properties, then we debar ourselves from commending that thing for having those properties. This is because commending something for being of a certain description becomes a tautology. For example, if being a good strawberry means ‘being large, red and juicy’ then in saying that ‘this strawberry is good because it is large, red and juicy’, we find ourselves saying that ‘this strawberry is large, red and juicy because it is large, red and juicy’. This, Hare argues, is what is wrong with attempting to reduce evaluation to description – not only do we end up making nonsensical remarks such as the one above, but, more importantly, we lose sight of the **evaluation** that we originally set out to make.
Secondly, if something’s being good were equivalent to that thing’s description, then there would exist infinitely many senses of ‘good’, each of which would have to be learnt individually, and none of which would convey what is commonly understood by ‘good’ regardless of what is being evaluated. For example, being ‘a good strawberry’ would mean being ‘large, red and juicy’, and being ‘a good novel’ would mean being ‘well phrased and interesting’, while being ‘a good long jumper’ would mean being ‘able to jump long distances’, and the list continues. None of these senses of ‘good’ have anything in common because ‘good’ is being used here to describe, not evaluate. However, because ‘good’ is a term of commendation, not of description, once learnt it can be generalised to all classes of things, and not equated with, or reduced to, the features of any.

Hare’s (1961) application of supervenience cannot, it follows, rely solely on the descriptive attributes (subvenient properties) of the thing or person judged. Rather, his claim is simply that in order for a person to be consistent, he or she cannot judge one thing to be good, whilst judging another identical thing to be non-good. It is in this sense that something’s being good is dependent on its descriptive properties (or in terms of supervenience, it is in this sense that ‘good’ supervenes on subvenient properties), however, the dependence is not on these properties alone.

Hare (1961) takes as an example the act of judging a man morally. In his words, “...the characteristics of the man (the minor or factual premiss) together with a specification of a standard for judging men morally (the major premiss), entails a moral judgement upon him” (pp.145-146).

What Hare has is a syllogistic argument. The first premise of the argument provides a standard for judging men. The second premise describes the man. The conclusion (the judgement of the man as ‘good’ or ‘non-good’) is entailed by the two premises. Thus, we can see that given the specification of the standard for judging men, it is entailed that two identical men must be judged such that it is concluded either that both are good men, or that neither is a good man. It is not simply the characteristics of the men that entail the judgement, it is their characteristics together with the standard for judgement.
The standard of judgement, therefore, is an essential component of the supervenience relation. So what are the characteristics of standards of judgement? According to Hare (1973), the standard of judgement in any act of evaluation is a universal prescriptive principle. He writes that, “To teach a person – or to decide on for oneself - a standard for judging the merits of objects of a certain class is to teach or decide on principles for choosing between objects of that class” (Hare, 1961, p.134). The universality of such principles means that they apply not only to actual objects that are judged, but also to hypothetical cases that are relevantly similar (Hare, 1984). Of course, where standards differ, so too will value judgements. Indeed, the same subvenient properties may validly be judged to determine a certain supervenient property under one principle, and the opposite supervenient property under a different principle. The existence of differing principles, even principles that are contrary to one another, need not be problematic if it is recognised that value judgements are made from a point of view. For Hare’s purposes, the claim that a major premise (i.e., standard) holds need not mean more than that it is subscribed to (Hare, 1984). Once a standard has been established (by whatever means standards are established), in other words, once a judge has subscribed to a standard of judgement for a particular class of objects, all objects that fall under the jurisdiction of the standard because of their descriptive properties can thereby be judged in accordance with the standard as to their value properties. Any instance of supervenience, Hare (1984) states, is therefore of this form:

\[
p: \text{For all } x, \text{ if } Gx \text{ then } Fx
\]

\[
q: G.a
\]

So \[
r: F.a
\]

In plain English, this argument reads: ‘For all individual cases (call them ‘x’), if x is G then x is F (p). This individual a is G (q), therefore this individual a is F (r).’

In neglecting the standard by which evaluative judgements are made, and in so doing focusing exclusively on subvenient and supervenient properties, philosophers sometimes misconstrue supervenience as a mysterious relation in need of explanation. Theories of reduction soon follow. However, the supervenience relation as used in ethics and aesthetics is neither mysterious nor amenable to reduction. In the context of making evaluative judgements, it is the person making the judgement, together with the
standard to which he or she adheres, that removes the mystery of why supervenient properties are linked to subvenient properties. Further, it is this dual dependence of supervenient properties upon both subvenient properties and the standard of judgement that precludes a naturalistic interpretation of supervenience. Reductive theses that ignore the role of the standard of judgement and thereby attempt to argue from the premise of a thing’s subvenient properties to the conclusion of a supervenient property lack a major premise, and are therefore invalid.

We are now in a position to understand the logical feature of supervenient properties that distinguishes them from other types of properties, including descriptive ones. It is the thesis that, “Necessarily, if \( r \), then there is a valid inference of the ‘\( p, q, \text{ so } r \)’ form, the two premisses of which hold” (Hare, 1984, p.5). This thesis states that wherever a supervenient property (\( r \)) exists, it is the case that there is a standard of judgement (\( p \)) and some combination of descriptive properties (\( q \)), which together entail the value property. The ‘necessarily’ here is important. The above form of argument can be used outside of the supervenience context, for example, to demonstrate a descriptive conclusion (such as in the ‘Socrates is mortal’ argument), however, the mortality of Socrates can be established independently of any form of argument (for example, by his dying), and therefore no argument of any kind is necessary for its establishment.

Supervenient properties such as value properties, on the other hand, necessarily involve the conjunction of a standard of judgement with descriptive properties, the combination of which they are consequences. There is no establishing a value property independently of an argument of the above form, and therefore of standards and descriptive properties. The consequential nature of supervenient properties is what distinguishes them from properties that ‘stand alone’.

The above sketch of supervenience raises an important issue concerning the ontological status of supervenient properties. Whereas a picture may have subvenient properties such as brightness and symmetry that exist independently of interaction with people, its supervenient properties such as goodness depend upon someone’s observing and evaluating its subvenient properties in the light of a standard of judgement. It is only when subvenient properties are evaluated that supervenient properties arise, and it is therefore only in particular contexts, those involving evaluators with standards, that supervenient properties can be said to exist. Supervenient properties, therefore, differ
from subvenient properties in that the former, but not the latter, are observer dependent. The observer dependence of supervenient properties is, of course, of no moment to ethical anti-realists such as Hare. Whether or not it is problematic in psychology remains to be seen.

2.5.2 Mental supervenience

Supervenience, as it has been characterised thus far, demonstrates both the dependence of supervenient properties upon subvenient properties (in conjunction with a standard of judgement) and the impossibility of ‘reducing’ (or giving a ‘naturalistic’ interpretation of) supervenient properties in terms of subvenient properties. Attracted to the non-reductive dependence that supervenient properties have upon subvenient properties in ethics, Davidson has sought to demonstrate an analogous situation in psychology, which may be called ‘mental supervenience’. By demonstrating that just as value properties supervene on descriptive properties, mental properties supervene on physical properties, Davidson hopes to prove that mental events are identical to physical events in an ontological sense (monism), but partially independent in a conceptual sense (anomalism). Through employing this distinction, Davidson aims at making sense of his assumptions that events are mental ‘only as described’, that mental events are simultaneously physical events, and that mental descriptions of events need not be related lawfully to physical descriptions of those same events.

2.5.2.1 Supervenient and subvenient properties in mental supervenience

The mental characteristics that Davidson takes to supervene on physical characteristics are the mental properties (or propositional attitudes) that are attributed to people through the process of radical interpretation. Less easy to identify are the physical characteristics upon which attributions of mental properties depend. Consider for example, “The definition of supervenience implies that a change in mental properties is always accompanied by a change in physical properties…” (1993a, p.7). Part of the vagueness here concerning which physical properties are being referred to is intentional, as seen in the continuation of the quotation, “… but it does not imply that the same physical properties change with the same mental properties” (p.7). If Davidson is correct, it is impossible to specify which physical properties are related to which mental properties, because there are no lawful connections between types of mental and physical properties. The vagueness, however, goes beyond omitting precise
specification of physical properties. We are left not knowing whether by ‘physical property’ Davidson means ‘of the person in question’s brain’ or ‘of the person’s movements and utterances’ or ‘of the person’s wider environment’, or some combination of these. This point, however, is essential for establishing Davidson’s conception of mental supervenience. If physical properties are to form the minor premise of the syllogism by which an observer attributes a belief or desire to a person, then clearly the physical properties must be observable. In the circumstances in which attributions are made, that is, in daily life, observers do not have access to the brains of others. Therefore, if it is brain states and events upon which Davidson deems mental properties to be dependent, his version of supervenience parts company with Hare’s.

There is some evidence that at times Davidson does think of the relevant subvenient properties as physical properties of the brain. Consider the following where a difference in a person’s intention is assumed to have its parallel in the brain, “There may, for example, be a difference in intention: this difference, we assume, has its physical aspect, since it is reflected in the propensities to motion of the agent. Given a complete description of the brain, we must expect this difference to correspond to some difference in physiology – ultimately in physics, as we have been seeing it” (1973c, p.252). Now, because an observer cannot access another’s brain, clearly a judgement concerning that other’s mental states could not be based on such information. Nor does Davidson think that the possession of such information would ever be of use in understanding another person’s mental states. He argues for quite the opposite, “If I am right, then, detailed knowledge of the physics or physiology of the brain, indeed of the whole of man, would not provide a shortcut to the kind of interpretation required for the application of sophisticated psychological concepts” (1973c, p.258). Either, then, Davidson has a very different notion of supervenience from that of Hare, one in which the existence of the supervenient property does not require observation of the subvenient properties, and one, therefore, in which the attribution of mental properties occurs quite independently of knowledge of the physical properties on which they are claimed to depend, or, alternatively, ‘physical properties’ must be understood more broadly. In addition to a person’s physiology, the broad perspective considers the movements and utterances that a person makes, and the environments in which they are made.
In a thought experiment, Davidson (1973c) asks his readers to imagine a robot named ‘Art’ that has been constructed so as to be physiologically indistinguishable from a human. Davidson’s aim is to consider whether complete knowledge of a person’s physiology would be of any use in understanding his or her psychology. The inclusion of this example here, however, is aimed at demonstrating that the practice of making psychological attributions relies not upon unobservable physical or physiological properties of the person interpreted, but on his or her observable physical characteristics, such as bodily movements and utterances. Davidson writes,

… let me turn back again for a moment to the question what makes us think Art has been properly constructed from a psychological point of view. I think the answer has to be, Art gives every appearance of thinking, acting, feeling like a man. And not just the superficial appearances. If you cut him he bleeds, if you shine lights in his eyes, he blinks, and if you dissect his eyes, you discover rods and cones. It is important, in deciding that he has psychological traits, that he is made like a man. If we found a radio receiver inside, and learned that another person was sending out signals to make Art move, we would no longer be tempted to assign psychological characteristics to Art. Any important differences under the skin might make us hesitate. Nevertheless, our detailed understanding of the physical workings cannot, in itself, force us to conclude that Art is angry, or that he believes Beethoven died in Vienna. In order to decide this, we would have first to observe Art’s macroscopic movements, and decide how to interpret them, in just the way we decide for humans” (p.250; emphasis added).

Even though Davidson makes mention of physical properties that are unobservable while discussing mental supervenience in an abstract way, when it comes to detailing with concrete examples how, exactly, a mental attribution is made, the physical properties involved are such things as bodily movements and utterances, together with the physical situations in which the movements and utterances are made. Nor should this come as a complete surprise. Radical interpretation, as we have seen, involves the process of ‘triangulation’ whereby the attribution of a mental characteristic, such as a belief, relies upon the interpreter observing both the responses of those they interpret, together with the environmental situations to which they respond. It is through taking note of the reactions (such as utterances) of others, while simultaneously observing
what it is that causes those reactions, that an interpreter is able to assign beliefs and other propositional attitudes to these objects of interpretation.

Indeed, the necessity of taking a broad or external view of the physical factors upon which mental attributions are made is used by Davidson as one of the arguments against the possibility of lawful connections between bodily and mental states. He claims that, “If mental properties are supervenient not only on the physical properties of the agent but in addition on the physical properties on the world outside the agent, there can be no hope of discovering laws that predict and explain behaviour solely on the basis of intrinsic features of agents” (1995b, p.5).

It thus appears that for the purposes of making mental attributions, knowledge of the specific functions of the brain is unnecessary. Ignorance of such information when making an attribution insures that it is not used, the irrelevance of such information insures that it could not be used.Attributions of mental descriptions require not simply the presence of certain physical states and events, whether of the persons interpreted or of the environmental situations in which they are located, but also the interpretation of these physical states and events within the relevant psychological context (as will soon be made clear). The subvenient properties in the mental supervenience relation are therefore understood by Davidson to be all of the physical properties of an individual’s body, actions and environment by which observers are able to judge whether or not that individual possesses a supervenient mental property. Thus far it appears that Davidson’s account of supervenience is consistent with Hare’s.

2.5.2.2 The standard in mental supervenience

As Hare has demonstrated in ethics and aesthetics, subvenient properties alone provide inadequate resources for an observer to make a judgement concerning the attribution of a supervenient property. What is required additionally for such attributions is a standard of judgement. Davidson, too, forsakes the notion of making attributions of mental properties based solely upon the presence or absence of individual physical properties such as utterances and bodily movements. Instead, as has already been indicated, psychological interpretation and the subsequent attribution of mental properties requires an holistic approach. This is evidenced in the by now well used quotation that, “It is a feature of the mental that the attribution of mental phenomena must be responsible to
the background of reasons, beliefs, and intentions of the individual” (1970a, p.222), and again by, “When we attribute a belief, a desire, a goal, an intention or a meaning to an agent, we necessarily operate within a system of concepts in part determined by the structure of beliefs and desires of the agent himself” (1974, p.80).

The attribution of supervenient mental properties to another may well depend upon the interpreter’s observation of that other’s subvenient physical properties, but such attributing need also take into account the background of interrelated attitudes with which the newly attributed characteristic must cohere. It follows, then, that a physical property, such as an utterance, will need to be interpreted with due consideration given to the speaker’s extant attitudes (i.e., those with which the speaker has been attributed already). Should these differ (from speaker to speaker, or within one speaker across time), so too will the interpretation of the physical property, and the subsequent mental attribution.

It thus seems that there is a clear functional parallel between the roles played by Hare’s ‘standard of judgement’ and Davidson’s ‘background of propositional attitudes’ in that both supplement subvenient properties and thus allow for a judgement or an attribution of a supervenient property to be made. However, this functional similarity between the two ‘standards’ obscures a fundamental difference in their characters. Hare’s standard is a universal principle connecting types of subvenient properties with types of supervenient properties. Accordingly, his version of supervenience is such that value judgements are entailed upon objects or persons that are subsumed under the principle. For example, in the following syllogism, the judgement or conclusion is lawfully deduced from the conjunction of the premises,

Major premise: All large, red and juicy strawberries are good
Minor premise: This is a large, red and juicy strawberry
Conclusion: This is a good strawberry

The character of Davidson’s ‘standard’, and also the procedure for making mental attributions, is more ambiguous. Conceivably, it could be that mental attributions are entailed by the background of propositional attitudes and the physical evidence that is interpreted, just as value properties are entailed by standards and subvenient descriptive
properties. This, however, if an accurate portrayal of the process of mental property attribution as envisaged by Davidson, reveals that rather than provide support for the theory of Anomalous Monism, supervenience brings about its undoing. As argued by Hare (1984), “As the link with universalizability shows, supervenience brings with it the claim that there is some ‘law’ which binds what supervenes to what it supervenes upon” (p.3). Furthermore, if the inference from subvenient to supervenient properties is to be valid, the law must be formulated in terms that explicitly mention both the subvenient and supervenient properties. In the strawberry example, the link between the descriptive and value properties is asserted in the universal premise. For Hare, this law connecting subvenient with supervenient properties is unproblematic, indeed necessary. The analogous situation in psychology, if the inference from subvenient physical properties to supervenient mental properties is to be valid, requires universal principles linking types of mental and physical properties. This means that there must exist laws connecting events described mentally with events described physically. Physical redescriptions of mental events cannot be substituted into the law on pain of invalidating the mentalistic conclusions that are drawn. Thus, any application of supervenience in psychology that is faithful to the ethical version requires psychophysical laws, and is therefore incompatible with mental anomalism.

Furthermore, Hare’s universal principle is fixed in a way that Davidson’s background of attitudes is not. For Hare, all descriptively similar items are to be evaluated alike because they are subsumed under an unvarying principle. Davidson, on the other hand, does not recognise any kind of priority for, of fixity of, the background attitudes by which subvenient properties are interpreted. Recall the idea of the ‘hermeneutic circle’ (Section I.iii) whereby interpretation of an individual’s actions (such as his or her movements and utterances) requires that his or her psychological make-up as a whole be consulted, and whereby, simultaneously, holistic understanding of that psychological make-up is influenced by the observation of individual actions. According to this notion, there is no strict priority granted to a pre-existing interpretation of an individual’s background of propositional attitudes, and thus it is possible that a new response by an individual, rather than necessarily being passively interpreted in line with the mental characteristics that have been attributed to him or her already, may result in some kind of adjustment of the overall picture. This is what Davidson is suggesting when he claims that, “… when we use the concepts of belief, desire, and the
rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory” (1970a, p.223).

The syllogistic form of Hare’s version of supervenience accounts for why descriptively similar objects and persons are consistently judged to possess similar value properties, at least in so far as a judge consistently applies the same standard of judgement. But, of course, Davidson hopes to employ supervenience to a different end, as a means of demonstrating the non-lawful dependence of mental on physical events. It is thus not open to Davidson to employ supervenience as Hare has done, because lawful entailments between events’ mental and physical descriptions are contrary to the thesis of mental anomaly. It appears, therefore, that while there are certainly parallels between Hare’s and Davidson’s usages of supervenience, ultimately Davidson deviates from the clear sense in which Hare originally employed the concept. In the latter’s opinion, if Davidson did mean to use ‘supervenience’ in a different sense, then, “… it might have been clearer if he had used another word, and, in telling us what he meant by it, had not appealed to the analogy” (Hare, 1984, p.16). Be that as it may, Davidson did appeal to the analogy, and he apparently had something at least similar to ethical supervenience in mind when he did so.

Davidson’s version of supervenience, like Hare’s, demonstrates that the act of attributing supervenient properties involves judging subvenient properties in line with other information. For Hare, this additional information entails a certain attribution. For Davidson, it provides a constraint on what attributions can be made. The constraint is such that an interpreter, who is aiming to attribute as rational a network of propositional attitudes as possible to another, must do so with due consideration given to which of the potential attributions best coheres with the network as a whole. However, this procedure is not lawful in the way that Hare’s entailments are, indeed, it may be the case that any of a number of differing mental properties are equally well suited for attribution to another, leaving the interpreter with no principled way of deciding between them. Davidson writes that, “… all possible evidence cannot limit acceptable theories to one. Given the richness of the structure represented by the set of one’s own sentences, and the nature of the connections between the members of this set and the world, we should
not be surprised if there are many ways of assigning our own sentences to the sentences and thoughts of someone else that capture everything of relevant significance… So we can say, if we please, that interpretation or translation is indeterminate, or that there is no fact of the matter as to what someone means by his or her words” (1996, pp.168-169). This ‘indeterminacy of interpretation’, as Davidson terms it, introduces a looseness to the procedure of making property attributions that is not present in Hare’s ethical supervenience. It is clear, therefore, that in adopting supervenience, Davidson did not mean to undermine Anomalous Monism by showing how physical and mental properties are, after all, lawfully connected. On what grounds, then, did Davidson believe that supervenience could be of use in resolving the paradox of Anomalous Monism?

2.5.2.3 The conceptual status of supervenient properties

To answer this question it is necessary to consider the ontological status of supervenient properties. As has already been indicated, Hare’s ethical supervenience is one in which value properties are not considered to be real features of the world that can be discovered through empirical investigation, rather, they are artefacts of what Andrew Melnyk (1991) terms our ‘projective tendencies’. Such tendencies are realised through our ascriptions of supervenient descriptions to objects and events in the world in systematic but ultimately erroneous ways. He writes that, “… while we ordinarily take our ascription of such properties to be true and accurate, we do so erroneously, since in reality there are no such properties. We project these non-existent properties onto the world as part of our attempt to explain it: the special sciences are ways of thinking of the world which aim to explain it, but the explanatory theories constituting them are false” (p.584).

The anti-realist underpinnings of supervenience suggest that supervenient properties have not an ontological but a logical or conceptual status. James Klagge (1988) has taken this to indicate that the judgements that give rise to supervenient properties need not reflect the nature of the world; instead, they reveal constraints such as conceptual requirements for the correct usage of the terms involved, “The necessity involved in ascriptive supervenience is conceptual necessity, since it derives from the meanings of

---

10 See Malpas (1989, 1992) and Searle (1987) for further discussion of Davidson’s ‘indeterminacy of interpretation’.
words” (p.464). Hare (1984), too, emphasises that attributions of supervenient properties are necessary because of logical rather than metaphysical considerations. He writes that, “In morals, likewise, some might say that universal moral principles hold necessarily; and others might prefer to say that they do not hold necessarily, but justify a necessary inference from themselves, plus a subsumptive premiss, to a moral judgement as conclusion… I prefer the second way of putting the matter” (p.11). The attribution of a supervenient property to an object or event is not, it follows, justified by the nature of the world but rather the conditions of usage for supervenient terms. The conditions, it was demonstrated in Section 1.5.1.2, are such that if a supervenient property applies to an object or event, then necessarily there has been exercised a judgement involving a universal principle and a minor premise by which the attribution of the supervenient property was logically entailed. It is this feature of supervenient properties that Hare (1961) aims to capture by calling them ‘consequential’.

Despite the proliferation of theories of mental supervenience that claim to provide a realist conception of mental properties, Davidson’s version is now well understood to belong to the anti-realist camp. Not always has this been the case, and as a consequence there exist several critiques of Davidson’s theory that, by assuming Anomalous Monism to be a realist theory, go wide of effectually showing its real weaknesses. This situation is not altogether surprising because far from declaring himself an anti-realist, Davidson has sought to defend his theory from such suggestions. He argues that, “Anomalous monism does not suggest that mental events and states are merely projected by the attributor onto an agent; on the contrary, it holds that mental events are as real as physical events, being identical with them, and attributions of states are as objective” (1997, p.112). This argument, however, is unlikely to stem the urgings of critics that Anomalous Monism is an anti-realist theory because it was never the reality of the states, events or acts of attribution that was called into doubt, rather, it was the reality of properties that require, for their existence, acts of attribution.

As occurs with value attributions, the objects and events to which mental properties are attributed are understood to have a mind-independent existence. Moral and mental properties, on the other hand, understood as arising out of acts of evaluation and interpretation, enjoy no such mind-independence. Why, then, does Davidson reject the claim that Anomalous Monism is an anti-realist theory, at least in so far as mental
properties are concerned? The answer, it turns out, is definitional. Returning to his analogy between numbers and the sentences used to attribute propositional attitudes, Davidson (1997) argues that numbers “… do not belong to the empirical world but to us, who need them in order to keep track of certain relations among objects” (p.114). Nevertheless, he likewise claims of numbers that they “… are as real as can be” (p.120).

It appears, then, that Davidson sees no contradiction in claiming that things not belonging to the empirical world are, nonetheless, real. Concerning this point, Davidson and his critics, myself included, differ in opinion. In claiming that Anomalous Monism is an anti-realist theory, we mean to indicate that it presupposes mental phenomena to belong not to the empirical world, but, like Hare’s value properties, as arising from our judgements concerning the empirical world. Our sense of something’s reality, therefore, is that it exists independently of us. Should something require our patronage for its continued existence, it is not real (see Section 4.1 for an introduction to realism).

By reminding us that, “…I have suggested that we take the actual utterance (or inscription) provided by the attributor as the object to which he is referring when [sic] to give the content of an attitude” (1997, p.115), Davidson is confirming that the content of an attitude is understood by him to arise out of acts of attribution whether spoken or written. Importantly, just as numbers are asserted not to be in objects, the content-giving sentences by which we attribute propositional attitudes “… are not in the thinkers – not even in their minds, or before their minds” (1997, p.114). Like the numbers, these sentences are not of the empirical world, but belong to us. By claiming, as he does, that, “Our sentences provide the only measure of the mental” (p.116), Davidson has adopted an anti-realist approach to the ‘mental’ components of mental events, where ‘realism’ is understood as indicating a thing’s mind-independent existence. Accordingly, while it may be granted that Davidson is a realist about states, events, and even acts of attribution themselves, it may not be granted similarly that he is a realist about mental properties. 

11 I specify ‘continued existence’ here in order to draw a distinction between the many real things, such as buildings, that require human patronage in order to come into existence (after which time they persist unaided), and the objects which are suggested to be without existence at the moment that they cease to be acknowledged, these latter being those I term ‘unreal’.

In accordance with the proffered characterisation of realism, there has been in recent years a growing appreciation of the importance to Anomalous Monism of maintaining a distinction between events (ontological items) and descriptions or property attributions (conceptual items), with the latter understood to be an artefact of the process of radical interpretation. Of particular interest has been the *a priori* maxim that radical interpretation should proceed so as best to construe others as rational beings. Clearly this pre-determined constraint is at odds with the open-minded approach to discovery that is more typically adopted in the investigation of nature. But it must be remembered that the practice of radical interpretation is not a means of investigating nature. As Louise Antony (1994) argues, “The justification for imputing rational structure to our actions is not empirical… rather it is *pragmatic*. And this makes the justification of psychological ascriptions ultimately instrumental: *if* we want to be able to view ourselves as rational agents, we must interpret ourselves so as to make applicable the principles of rationality” (p.235).

It is not, according to Davidson, an open question whether or not an individual’s behaviour is rational because the concept of behaviour is *defined* in terms of rational connections between thought, desire and action. In the absence of these connections, something other than behaviour is occurring, by definition. Behaviour, alongside desire and thought, is a conceptual notion. It is not an independent fact of nature, but a description one imposes upon another with a view to understanding and explaining that other from a psychological point of view. In so doing, one does not simply accept whatever evidence becomes available, rather, “As long as it is behavior and not something else we want to explain and describe, we must warp the evidence to fit this frame” (1974, p.85). As Klagge (1990) sums up this situation, “We ascribe beliefs and desires to people, in part, as a way of understanding, predicting, and appraising their behavior. Thus, the mental becomes more a way of seeing people than it is something in people that can be seen” (p.342).

To return, then, to the question of why Davidson might have thought that supervenience could be of use to the theory of Anomalous Monism, it is instructive to see that supervenience allows for the conception of a realm of properties that are constituted not by the facts of nature, but by the descriptions imposed upon nature from the point of view of a judge or interpreter, “The notion of supervenience, as I have used it, is best
thought of as a relation between a predicate and a set of predicates in a language” (1985d, p.242). This allows for the required division between events themselves, which are independently existing ontological items or facts of nature, and event descriptions, which are comprised of properties or predicates that have been attributed to the events according to a set of principles or other predetermined conceptual constraints. By proposing discrepant ontological statuses for events and their descriptions, Davidson has constructed a picture of the world whereby events are real but properties are artefacts of interpretation. Because of its reliance on the interpretations of humans, the classification of events according to types of properties is subjective, tentative, and, of course, without basis in the world itself. The role of supervenience in Davidson’s theory is therefore one of articulating the intuition of Anomalous Monism that while events and causal relations are mind-independent, classifications and explanations of events are constitutively conceptual.

2.6 Resolving the paradox

By proposing the theory of Anomalous Monism, Davidson aims to resolve the paradox implicit within the view that, despite the lawful character of causality and the absence of psychological laws, mental events are nevertheless causally efficacious. The key to resolving the paradox, Davidson believes, is to recognise that it is events, not properties as described in one way or another, that enter into causal relations, and thus that whether or not an event is efficacious is independent of the fact of whether or not its various descriptions are appropriately formulated for inclusion in laws.

Nevertheless, it is by no means intuitively clear how it can be that events, but not their described properties, are involved in causal relations. It is the clarification of this matter that Davidson aims to achieve by proposing discrepant ontological statuses for events and their properties. He argues that events, like objects, are mind-independent entities in the world, but that properties, descriptions, classifications and explanations of events are conceptual, that is, dependent upon the interpretational practices of humans. The necessary divorce between the facts of events’ existences and relations on one hand, and the descriptions of these on the other, is captured, Davidson believes, by the position of Anomalous Monism. This, he has recently reminded his readers, is “… the position that says there are no strictly lawlike correlations between phenomena classified as mental and phenomena classified as physical, though mental entities are identical, taken one at
a time, with physical entities. In other words, there is a single ontology, but more than one way of describing and explaining the items in the ontology” (1995b, p.4).

The proposed ontological-conceptual distinction between events and their property descriptions is aided by the hypothesis that there exists a relation of supervenience between the mental and the physical properties of an event. If it is sound, mental supervenience might provide Davidson with what he needs to reconcile his doctrine of free human action with the nomological character of causality in general. As he expresses it, “… the ontological reduction, if it succeeds, is enough to answer many puzzles about the relation between the mind and the body, and to explain the possibility of autonomous action in a world of causality” (1978, p.88). This is because autonomous activity, that is, activity that is portrayed by means of the concepts of folk psychology, will then be understood to be autonomous only in so far as it is described appropriately. With the discrepancy between what an event is (ontology), and how an event is described (conceptuality), firmly in mind, it remains only to remember that the same events that are described with the concepts of folk psychology can be redescribed in the language of physics and, correspondingly, that free actions are simultaneously strictly determined events, “… it is only events as described in the vocabulary of thought and action that resist incorporation into a closed deterministic system. These same events, described in appropriate physical terms, are as amenable to prediction and explanation as any” (1974, p.80).

Anomalous Monism, if true, allows the following:

1. Monism: taken individually, mental events are identical with physical events. Of course, this formulation is a little misleading. It should not be understood to suggest the identity of two events, rather, the amenability of one event to being described in more than one way.

2. Mental anomalism: the classification of an event as belonging to a mental type proceeds without lawful connection to the classification of that same event as a physical type. Furthermore, mental classification is tentative. It alters accordingly with the variety of principles upheld by the individuals making the classification, as well as evolving over time.

3. The causal relevance of mental events: causal relations obtain between individual events independently of those events’ various classifications. Some
events that act as causes and effects have mental descriptions and it thus follows that mental events are involved in causal relations.

4. Causal nomologicality: the fact that anomalous mental events are causally efficacious provides no evidence against the claim that all instances of causation are nomological because every anomalous event can be redescribed in alternative physical terms that are amenable to inclusion in strict laws. This accords with the principle of the nomological character of causality, which Davidson asserts is to be read to say, “… when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law” (1970a, p.215).

One can thus conclude, along with Davidson, that what gave the appearance of a paradox was in actuality a confusion that arose from a failure to observe the ontological-conceptual distinction between events themselves and the descriptions attributed to events. However, in order to be satisfied with this solution, one must also adopt an anti-realist stance that relegates mental properties to the status of being the mind-dependent conceptual artefacts of third-person interpreters. The tenability of the theory of Anomalous Monism will be considered in Section 3.
Section 3: Criticism of Davidson’s Anomalous Monism

Criticism of Davidson’s theory of Anomalous Monism supplemented by weak mental supervenience abounds in the secondary literature. Strangely, the essence of Davidson’s theory, that ‘events are mental only as described’, is oftentimes overlooked, and in these cases arguments against the theory fall short of refuting it in that they do not address it properly. This is not to say that the criticisms have no value, rather that their value is undermined because they assume, in a way that Davidson does not, mental realism. As will be argued, it is the anti-realist assumption that events’ properties are artefacts of human linguistic practices that is the real culprit within Davidson’s theory. From this starting point, the view has arisen that, in and of themselves, independently of descriptions, events do not have properties. The assumed propertyless characters of events has allowed a counterintuitive division between causal relations and the laws of which they are instantiations to develop, and it is this that has allowed Davidson to maintain that mental phenomena are both anomalous and causally efficacious, despite causation being intimately tied in with strict causal laws.

In accordance with the above comments, it is acknowledged that powerful criticisms have been developed concerning the capacity of Anomalous Monism to provide the key to mental causation and explanation, but it is also suggested that these criticisms are themselves superfluous because they treat Anomalous Monism as if it were a viable theory of the character of mental phenomena, and only from this false beginning do they demonstrate that it is nevertheless incapable of providing a mechanism either for mental causation or the related field of causal explanation. However, as will be argued, Anomalous Monism is not a viable theory of mental phenomena. It is the main aim of the present section to establish that the fundamental anti-realist assumptions upon which Anomalous Monism is founded require that it be rejected outrightly, and, accordingly, that the proposed solution to the paradox of mental causation that pivots upon the theory be similarly discarded.

In the present section, five criticisms of Davidson’s theory of Anomalous Monism and weak mental supervenience are to be considered. The first three of these are arguments that deal with some of the alleged consequences of Davidson’s theory of the character of mental phenomena when it is taken to be a viable one, namely internal inconsistency (Section 3.1), mental epiphenomenalism (Section 3.2) and explanatory failure (Section
3.3). The subsequent two criticisms concern the fundamental anti-realist tenets within Davidson’s theory, that mental properties are linguistic artefacts (Section 3.4) and that events are propertyless particulars (Section 3.5).

3.1 Weak mental supervenience, ‘possible worlds’ and internal inconsistency

Davidson’s usage of mental supervenience as a means of characterising the relationship between mental and physical properties has attracted much attention and criticism. To a large extent, the criticism levelled at Davidson has arisen because of his failure to spell out precisely the mechanism by which supervenience works. Due to its importance within the theory of Anomalous Monism, this is an unfortunate omission that has led to much unnecessary confusion as demonstrated below.

Of primary concern to some of Davidson’s critics is the view that his ‘weak’ version of mental supervenience is unable to capture the notion of dependence between mental and physical properties that is necessary for a monistic theory of mind. Consequently, it is believed that mental supervenience, as employed by Davidson, is internally inconsistent, as can be demonstrated by the use of ‘possible worlds’.

The use of ‘possible worlds’ in this literature is an heuristic device employed in thought experiments. Possible worlds can be imagined to be the same as the actual world, differing from it in some details, or of a completely different character. Most typically in the mental supervenience context, possible worlds are imagined to be identical to the actual world in all physical details. However, whether they are also imagined to be identical in all mental details depends on the kind of supervenience under investigation. In fact, this distinction delineates the differing senses of ‘dependence’ between strong and weak supervenience.

Theories of strong supervenience claim that “…base properties must determine supervenient properties in the sense that once the former are fixed for an object, there is no freedom to vary the latter for that object (Kim, 1984a, p.60). In other words, according to theories of strong supervenience, the presence of certain combinations of subvenient physical properties is sufficient for determining supervenient mental properties. There is no possibility of two individuals being physically identical yet
mentally distinguishable. In terms of possible worlds, if the actual world and a possible world contain individuals who are physically identical, then those individuals will necessarily be psychologically identical too. Recalling the indiscernibility requirement from Section 2.5.1.1, whereby objects that are indiscernible in their subvenient properties must be indiscernible in their supervenient properties, it follows that strong supervenience is taken to be the indiscernibility requirement with respect to all possible worlds.

In cases of weak supervenience, such as Davidson’s, it has often been taken to be the case that the supervenient properties are dependent upon the subvenient properties within this world, but that this relationship does not necessarily hold across all possible worlds, “The weakness of weak supervenience lies in this: as defined it prohibits the possibility of two things agreeing in base properties but differing in supervenient properties only within any possible world” (Rowlands, 1995, p.10). Thus, according to theories of weak supervenience, it is possible that a certain combination of subvenient properties will consistently underlie a particular supervenient property within this world, while, in another world, the same combination of subvenient properties could underlie a different supervenient property. Therefore, weak supervenience is the indiscernibility requirement with respect to one world, the actual world, only.

That Davidson’s version of mental supervenience conforms to this pattern is assumed to be shown by his claim that “… there cannot be two events alike in all physical respects but differing in some mental respect…” (1970a, p.214), together with his insistence that there are no strict laws linking mental and physical properties. This combination of claims suggests that while it is, in fact, in the actual world, the case that a particular combination of physical properties determines a particular mental property, things might well have been different, and, therefore, that there is no impediment to imagining that they are different, in other possible worlds.

Understanding, then, that ‘weak supervenience’ is taken in the secondary literature to be the indiscernibility requirement with respect to one world, it is possible to discern how the theory gives an impression of internal inconsistency. Simon Blackburn (1985) illustrates the situation as follows:
In any possible world, once there is a thing which is F, and whose F-ness is underlain by G*, then anything else which is G* is F as well. However, there are possible worlds in which things are G* but not F. Call the former worlds G*/F worlds, and the latter, G*/O worlds. The one thing we do not have is any mixed world, where some things are G* and F, and some are G* but not F… My form of problem, or mystery, now begins to appear. Why should the possible worlds partition into only the two kinds, and not into the three kinds? (p.53)

Blackburn’s puzzle is a fairly straightforward one. If within one world, call it W1, a combination of subvenient physical properties (G*) is sufficient for determining a supervenient mental property (F), and if within a second world, call it W2, the same combination of subvenient physical properties (G*) is insufficient for determining that mental property (hence underlying O), then what is to prevent inter-world travel whereby an individual from W2 who possesses the combination of physical properties (G*) but not the mental property (O) comes to reside in W1? If there is nothing preventing inter-world travel, then it is possible that W1 could contain physically indistinguishable individuals who differed mentally, which is to say that W1 could be a ‘mixed world’. However, if indeed weak supervenience is the indiscernibility thesis with respect to one world, such an outcome should be impossible. So what has gone wrong?

According to Rowlands (1995), “The problem with weak supervenience is not that it is a weak relation of determination; the problem is that it is not a relation of determination at all” (p.11). It is easy to see how such a conclusion has been reached. Having decided that theories of weak supervenience, such as Davidson’s, allow discrepant mental properties to supervene on identical physical properties across worlds whilst assuming that, within any one world, a particular combination of subvenient properties consistently determines a particular supervenient property, critics such as Rowlands are inevitably brought to the conclusion that, in cases of weak supervenience, subvenient properties do not determine supervenient properties at all. How could they when mixed worlds are possible? Weak supervenience, it appears, is internally inconsistent in that it both insists that subvenient properties determine supervenient properties (within one world), and simultaneously allows that this is not the case. The solution, Rowlands claims, is that weak supervenience must be ‘strengthened’.
However, both strong supervenience, and the reading of Davidson’s mental supervenience as the indiscernibility requirement within one world, are at odds with the notion of supervenience itself. Strong mental supervenience, whereby supervenient mental properties have some kind of necessary connection with subvenient physical properties, is a version of the naturalist theory of supervenience that Hare (1961) dismisses as unworkable (see Section 2.5.1.2). The problem with asserting that subvenient properties alone determine supervenient properties is that the role played by the standard (or major premise) is ignored.

Similarly, according to Davidson’s version of mental supervenience, the attribution of supervenient mental properties requires both observation of physical properties and interpretation of those properties in accordance with a background of propositional attitudes. The mere presence of the physical properties alone is insufficient for the assigning of mental content, whether in this or any other world. Indeed, it does not require another world for the same set of subvenient properties to determine differing supervenient properties, rather, all that is required are two differing standards (or in the case of mental supervenience, two different backgrounds of propositional attitudes). Mixed worlds, it seems, are no threat whatsoever to the consistency of Davidson’s mental supervenience. Nor should they come as a surprise. As has been repeatedly emphasised above, it is because there are no strict laws between mental properties and the physical properties that they supervene upon that physically indistinguishable individuals may differ mentally, within this world. It is therefore a misunderstanding to claim that Davidson’s version of supervenience is defined by the indiscernibility requirement within one world, and consequently a mistake to claim that it needs to be in some sense strengthened in order to become internally consistent.

Despite the ease with which the apparent problem of mixed worlds can be swept away, it is not difficult to see how the standard in supervenience came to be ignored, and thus how the appearance of inconsistency first arose. It is Davidson’s own definition of supervenience, the one contained in his oft-quoted article, that has done the mischief. In that article, Davidson (1970a) claims that, “… supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect…” (p.214). He makes no explicit mention at this or at any other point in his writings of what else, namely a standard, is required for attributing supervenient
mental properties. Indeed, it is only through both acquainting oneself with supervenience in its original context, and through noticing the role played by the background of propositional attitudes in Davidson’s system of mental attributions, that the omission can be detected.

It must therefore be concluded that Davidson’s version of mental supervenience is internally consistent. There remains the more fundamental question, nevertheless, of whether it is appropriate to view mental characteristics as supervenient upon physical characteristics at all. This question is taken up in Section 3.4.2 below.

3.2 Anomalous Monism, mental supervenience and mental epiphenomenalism

One of the most persistent criticisms of Davidson’s theory of Anomalous Monism, supplemented by his version of mental supervenience, is that it is consistent with, and perhaps entails, mental epiphenomenalism. Of course, given that a central aim of Davidson’s theory is to demonstrate that mental events cause and are caused by other events, whether mental or physical, the charge of mental epiphenomenalism is an important one. Davidson’s reply to the allegation that Anomalous Monism leads to mental inertness revolves around the distinction he draws between extensional causal relations and intensional causal explanations, including laws. It is the conflation of these, he argues, that has confused his critics. Unlike them, Davidson believes that Anomalous Monism provides the basis for a plausible theory of mental causation.

The paradox with which Section 1 closed was that Davidson’s three principles of Anomalous Monism (Causal Interaction, Causal Nomologicality and Mental Anomaly) appear to give rise to an inconsistent triad. The problem, it seemed, was that while all causal relations are subsumed under strict deterministic laws, mental events are anomalous, that is, free from subsumption under strict laws, and yet that mental events are causally efficacious. Consequently, it appeared that at least one of Davidson’s principles was false. Nevertheless, as was spelt out in Section 2, it is possible to maintain all three of Davidson’s principles if events are ontological particulars while

---

explanations, including laws, are descriptive sentences formulated in accordance with the attribution of properties. Or at least so it has seemed to Davidson. His critics, on the other hand, while accepting, to some extent, that events are particulars and that causation is an extensional relation between individual events, have nevertheless sought to prove that “The fact is that under Davidson’s anomalous monism, mentality does no causal work” (Kim, 1989b, p.35).

The arguments proposing that if mental phenomena are anomalous, then they are causally inefficacious, come in different varieties, of which two will be considered. The first of these is that mental events are causally inert. The second is that mental events are efficacious, but only because they are simultaneously physical events, their mental properties being irrelevant. The distinction between the two varieties of argument will readily become clear.

An early proponent of the view that Davidson’s Anomalous Monism renders mental events causally inert was Kim (1984b). He read Davidson’s thesis as saying that “… whether or not a given event has a mental description (optional reading: whether it has a mental characteristic) seems entirely irrelevant to what causal relations it enters into. Its causal powers are wholly determined by the physical description or characteristic that holds for it; for it is under its physical description that it may be subsumed under a causal law. And Davidson explicitly denies any possibility of a nomological connection between an event’s mental description and its physical description that could bring the mental into the causal picture” (p.267).

It is not surprising that Davidson should react to criticism such as this with amazement. He does not claim that an event’s causal powers are wholly determined by its physical description, indeed, he takes pains to stress that how an event is described is irrelevant to its causal relations. The fact, as Davidson sees it, that only physical descriptions of events are amenable to refinement into strict laws can in no way undermine either the utility of other causal explanations (that is, those using alternative descriptions, including mentalistic descriptions) or, more importantly, an event’s capacity to act as a cause. It is presumably with criticism of this kind in mind that Davidson (1993a) reiterates that, “We can say, if we please (though I do not think this is a happy way of putting the point), that events instantiate a law only as described in one way rather than
another, but we cannot say that an event caused another only as described. Redescribing an event cannot change what it causes, or change the event’s causal efficacy” (pp.6-7).

Here we see Davidson’s appeal to the distinction between the extensional character of events and causal relations (that is, to the view that causal relations obtain between events independently of how they are described), and the intensional character of explanations (which depends explicitly upon how events are described). Because causal efficacy relates only to causal relations themselves, how an event is described, that is, mentally or physically, can be of no concern to whether or not that event is efficacious. Consequently, the accusation of mental event epiphenomenalism under Anomalous Monism fails in its present form because it assumes, contrary to Davidson’s writings, that events draw their causal efficacy from their physical descriptions.

The second variety of mental epiphenomenalism of which Davidson is accused concerns not mental events, but mental properties. Brian McLaughlin (1993) readily absolves Davidson of the charge of mental event epiphenomenalism and asserts that none of his critics really believe him to be guilty of that. To this end he writes that, “When they charge that he is committed to epiphenomenalism, they do not mean that he is committed to denying that mental events cause physical events. They mean, rather, that he is committed to the view that when mental events cause physical events, they do so in virtue of falling under physical types, and not in virtue of falling under mental types” (p.27). Indeed, it seems clear from other of Kim’s writings that his suggestion of mentality doing no causal work, under Anomalous Monism, is meant to be understood in this second sense, “…by ‘mentality’ I was referring to mental properties, not individual mental events” (Kim, 1993b, p.20), and Davidson (1993a) seems also to recognise that what his critics have in mind is the question of whether “… the mental properties of an event make no difference to its causal relations?” (p.13).

The view that Anomalous Monism entails mental property epiphenomenalism turns on the assumption that physical causation comprises a closed system, one in which the physical properties of causal events alone suffice to bring about effects. For example, in terms of the causation of a behaviour such as walking to the bus stop, it can be argued that there exist necessary and sufficient physiological states and events that bring about the appropriate bodily movements, and therefore that any mental properties that happen to accompany these physical properties are superfluous. The situation appears to be
even more dire than this for Anomalous Monism, because the unlawful connection it proposes to exist between mental and physical properties prohibits there existing so much as a correlation between them, one that might suggest that mental properties constrain physical properties and therefore are relevant to physical causation in an indirect way. Consequently, it has appeared to its critics that Anomalous Monism conjoined with weak supervenience entails mental property epiphenomenalism because, “… the very same network of causal relations would obtain in Davidson’s world if you were to redistribute mental properties over its events any way you like; you would not disturb a single causal relation if you randomly and arbitrarily reassigned mental properties to events, or even removed mentality entirely from the world” (Kim, 1989b, pp.34-35). In other words, its critics have concluded that Anomalous Monism, supplemented by weak supervenience, dictates that events enter into causal relations indifferently to the presence or absence of mental properties.

Before pursuing this line of reasoning too far, it is necessary to pause and reconsider what is being proposed by both Davidson and his critics, and thereby to identify the source of what has become a debate of considerable duration, little progress, and much arguing at cross purposes.

In short, the critics’ concern is that Davidson’s thesis entails that mental properties are causally irrelevant and it is Davidson’s claim that how events are described, and therefore the properties that are attributed to them, has no impact upon an event’s causal efficacy. Well may it be asked, wherein lies the dispute? Why cannot it be agreed by all that Anomalous Monism does entail mental property epiphenomenalism because it denies mental properties an ontological status, whilst maintaining that causation is an ontological relation? If this were to happen, and Davidson’s conception of causation as an extensional relation were accepted, then it could still be maintained that Anomalous Monism is an acceptable theory of mental causation because it would be understood that it is events, not their properties, that enter into causal relations.

This proposed solution, however, will not do. The charge of mental property epiphenomenalism is typically deemed to be a more momentous criticism of Anomalous Monism than is the charge of mental event epiphenomenalism, precisely because most authors are agreed that events derive their efficacy from their properties,
and thus that the question of causal relevance is one best directed at properties, not events. For example, Jackson and Pettit (1990) comment that, “In addition to asking which events are causally relevant to which other events, we can and must ask which properties of events are causally relevant to which other properties. But surely this must, or should, have been an implicit ingredient in the story about causation all along. Surely not even the most robust defender of a concrete conception of events supposed that featureless events might do some causing. Their events caused what they did because of how they were – that is to say, because of which properties they possessed” (p.197).

What Jackson and Pettit (1990) articulate here is easily discerned as an often tacit assumption that is found in much of the criticism aimed at Anomalous Monism, that events gain their causal efficacy from their properties, and hence that a conviction of mental property epiphenomenalism is, on its own, sufficient to undermine Anomalous Monism’s claim of being a plausible theory of mental causation. I shall refer to this assumption as the ‘property assumption’.

Note that as soon as the focus of discussion shifts like this from the causal relevance of events to that of their properties, an appreciation that the property assumption is generally presupposed becomes crucial, because it is an assumption with which Davidson does not concur. Given that this discrepancy between Davidson and the majority of authors exists, it is surprising to find that it is rarely mentioned by the latter. Quite possibly this neglect occurs because Davidson’s position has been deemed ambiguous, or, alternatively, because it has simply been misunderstood.

For example, Kim (1993b) suggests that, “… if something that purports to be a theory of mental causation assigns no causal role to mental properties – if it has nothing to say about the causal powers of mental properties while saying plenty about those of physical properties – the theory can, it seems to me, reasonably be said to be epiphenomenalistic with regard to mental properties” (pp.20-21). Kim’s argument turns on the claim that Davidson has plenty to say about the causal powers of physical properties, yet nothing to say about the causal powers of mental properties. And, indeed, if it could be shown that Davidson’s theory does rest on the efficacy of physical properties and, presumably, the supervenient dependence of mental upon physical
properties, then Kim is right to argue that it is, at best, only in some derived sense that mental properties could be causally relevant.

A question thus arises concerning why Kim supposes that Davidson agrees that an event’s physical properties are relevant to its causal efficacy, when Davidson has repeatedly claimed that whether an event has a certain property or not depends entirely upon how it is described, whereas its causal efficacy obtains independently of how it is described. One need not search far to find the answer. The paper in which Kim (1993b) suggests that Davidson has plenty to say about the causal powers of physical properties was written as a response to a paper of Davidson’s that contains claims such as, “If supervenience holds, psychological properties make a difference to the causal relations of an event, for they matter to the physical properties, and the physical properties matter to causal relations” (1993a, p.14). A remark such as this seems to justify Kim’s views that, firstly, Davidson believes physical properties to be relevant to an event’s causal efficacy and, secondly, that if mental properties are causally relevant, it is because they are superveniently linked to physical properties. It is a short path from these admissions to the view that an event’s physical but not mental properties are the relevant factors in its causal relations, and thus indeed that while ‘mentality does no causal work’, physicality does plenty.

The ambiguity that arises between Davidson’s argument, just mentioned, for the relevance of properties to an event’s causal capacity and his view that properties are only descriptions which have no bearing upon the causal efficacy of the events to which they are attributed is one that cannot be resolved easily. Of course, it could be that Davidson has changed his mind about the relevance of property descriptions to an event’s causal efficacy, but contained within the same paper that asserts that properties do make a difference to an event’s causal relations is a confirmation that, “… it is also irrelevant to the causal efficacy of physical events that they can be described in the physical vocabulary. It is events that have the power to change things, not our various ways of describing them” (1993a, p.12), and again, “Given this extensionalist view of causal relations, it makes no literal sense, as I remarked above, to speak of an event causing something as mental, or by virtue of its mental properties, or as described in one way or another” (1993a, p.13). It appears, then, that Davidson has not changed his mind
about the relevance, or rather, irrelevance, of an event’s ascribed properties to its causal efficacy.

Is there another way of resolving the ambiguity? Is it possible to maintain an extensionalist view of causation whilst recognising the causal relevance of an event’s properties? In other words, is it possible to reconcile Davidson’s claims because an extensionalist view of causal relations is consistent with the view that when an event is causally efficacious, it is so in virtue of its properties? A reconciliation of this sort is recommended by McLaughlin (1993).

McLaughlin (1993) understands that, according to Davidson, to hold that causal relations are extensional is to claim that “… it makes no literal sense to speak of an event causing something in virtue of any of its properties” (p.30), but he believes that Davidson’s claim is mistaken. It is possible, according to McLaughlin, to hold an extensionalist view of causal relations and simultaneously to hold that if an event is causally efficacious, then it is so in virtue of its properties. In support of this view, McLaughlin cites an analogous example, that of the ‘weighs-less-than’ relation. This, too, he argues, is an extensional relation between non-abstract particulars, and one in which it makes sense to say that “… if a weighs less than b, then this is so in virtue of something about each, namely, their weights” (p.31). He elaborates, “If Tom weighs more than Mary, then they are so related however they are described. None the less, if Tom weighs more than Mary, then they are so related in virtue of the fact that Tom has some weight, \(w_1\), Mary, some weight \(w_2\), and \(w_2\) is less than \(w_1\)” (p.32). In a similar vein McLaughlin (1993) argues, “That causal relations are extensional relations between events is straightforwardly compatible with the claim that when events are causally related, they are so in virtue of something about each” (p.32). The ‘something’ in virtue of which events are causally related is, according to McLaughlin, their properties.

Notice, however, that McLaughlin’s argument rests on a realist conception of properties that Davidson does not share. By arguing that if Tom weighs more than Mary, then he does so in virtue of their respective weights, that is, because of their inherent physical properties, McLaughlin assumes that objects and events really have properties, rather than that they are simply spoken of in that way. Davidson, it will be recalled from Section 2.2.2, thinks very differently. He argues that, “In thinking and talking of the
weights of physical objects we do not need to suppose there are such things as weights for objects to have. Similarly in thinking and talking about the beliefs of people we needn’t suppose there are such entities as beliefs” (1989b, p.11). What McLaughlin has ignored is that, when Davidson speaks of an object or event as being ‘particular’, he means more than simply that each object and event is uniquely located in time and space. In addition to this, Davidson means that, in and of themselves, objects and events are devoid of properties. It is only through human conceptualisations that objects and events come to be thought of, spoken about, and treated as if they were propertied. Literally speaking, on Davidson’s view, events do not have properties. Similarly, by calling causal relations ‘extensional’, Davidson is emphasising that they proceed indifferently to any properties that are attributed to the events involved in them. McLaughlin is therefore mistaken to think that an extensionalist view of causal relations, such as Davidson’s, is compatible with the causal efficacy of properties. Such a combination would be self-contradictory.

Therefore, despite the urgings of Jackson and Pettit (1990), it must be admitted that some authors do indeed suppose that featureless, or at least propertyless, events are causally efficacious, and that prominent among these is Davidson. A refusal to admit just this appears to be at the heart of the drawn out debate between Davidson and his critics over whether Anomalous Monism does, or does not, entail mental epiphenomenalism. As Davidson has repeated again and again, Anomalous Monism takes it for granted that properties are varieties of description, and that an event’s being efficacious occurs indifferently to how it is described. Consequently, what appears plain to Jackson and Pettit, the platitude that events are efficacious in virtue of their properties, needs to be demonstrated if it is to have the desired impact of revealing that Anomalous Monism entails mental epiphenomenalism.

Finally, there remains the question why, if Davidson believes events to be devoid of properties, he would attempt to defend his theory against a charge of mental property epiphenomenalism, and why, in particular, he would do so by making a claim concerning the role of properties in events’ causal relations. Perhaps he has mistakenly believed that accepting the charge of mental property epiphenomenalism, which should be of no concern to one who believes that properties are mere words used in explanations, would subsequently justify the charge of mental event epiphenomenalism.
which, if his views concerning the ontological distinction between properties and events are correct, it would not. Or perhaps he concurs with Norman Melchert (1986) and Tim Crane (1992) that the charge of mental property epiphenomenalism is misdirected at Anomalous Monism from the start because, “The mental properties of events are not left dangling inefficaciously, for the reason that there aren’t any distinctive mental properties in the world” (Melchert, 1986, p.274), and “… to formulate the charge properly, you have to be able to ask: in virtue of which property of the cause does it have its effect? But this is a question that Davidson’s nominalist ontology does not let you ask” (Crane, 1992, p.189). If these speculations or something like them cannot account for Davidson’s concern over Anomalous Monism’s purported mental property epiphenomenalism, then it is unclear, given his extensionalist view of events, what can.

In summary, when the proposed separation of events and properties into ontological and conceptual categories is left unchallenged, Anomalous Monism remains immune to the charge of mental event epiphenomenalism, and indifferent to the charge of mental property epiphenomenalism. Given that this indifference is justified when an extensional theory of causation is accepted, or at least unchallenged, it remains a matter of conjecture why Davidson sees fit to oppose the charge of mental property epiphenomenalism.

3.3 Anomalous Monism and explanatory failure

Even if, as Davidson maintains, mental events are causally efficacious, there remains the concern that rationalisations, that is, causal explanations of actions in terms of reasons (i.e., desires and beliefs), are incompatible with the anomalous character of mental phenomena. Recall that, according to Anomalous Monism, there are no strict deterministic laws by which events, when described mentally, can be predicted and explained. When mental events enter into causal relations there is always an applicable strict law at hand, but this strict law is formulated not in terms of the events’ mental descriptions, rather, their physical descriptions. It thus happens that because of their anomalous character, mental events, so described, are at best amenable to inclusion only in ‘low-grade’ explanations involving rough generalisations and poor probabilities. Yet the very idea of explanation seems to require something more than this. It seems to require lawful connections between what is to be explained (the explanandum) and what is to do the explaining (the explanans). Before investigating these matters, it is first
necessary to define and distinguish the terms ‘universal’ and ‘general’ as they apply to laws. The importance of observing the distinction between these two concepts will readily become clear.

The ‘universality’ of a principle or law means that it is applicable to all members of a specified class. For example, the law that ‘all humans are animals’ is universal because it states that all members of a certain class, humans, are of a certain description, animalistic. Similarly, the law that ‘all mammals are animals’ is universal because it states that all members of a certain class, mammals, are of a certain description, animalistic. Neither statement is more or less universal than the other, indeed, it makes no literal sense to suggest that there could be degrees of universality (see Hare, 1973, p.6). It does, however, make sense to suggest that ‘all mammals are animals’ is more general than ‘all humans are animals’ because the class to which it refers, ‘mammals’, is less specific than the class ‘humans’. In this way, universal laws can differ from each other as to their degree of generality. To continue the example, ‘all female humans are animals’ is more specific, and hence less general than the laws hitherto mentioned, while ‘all vertebrates are animals’ is less specific, and hence more general.

With the distinction between universality and generality in mind, it is possible to consider the view that lawful connections between explananda and their explanans are necessary for explanations. One such argument has been put forward by R.C. Solomon (1974), who aims to defend the view that reason explanation is a species of causal explanation. Solomon argues that the very notion of explanation presupposes laws, and therefore that “… some general principle is necessary for any explanation of an action by citing reasons, at least the very general principle that motives and beliefs together determine action” (p.424).

Solomon’s (1974) argument for the necessity of a general universal principle connecting explananda with their explanans bears a great deal of resemblance to Hare’s (1961) argument concerning the necessity for a standard (or major premise) in the supervenience relation according to which subvenient properties may be judged, and supervenient properties attributed (see Section 2.5.1.2). Like supervenience, explanation is deemed to take the form of a syllogistic argument consisting of two premises and a conclusion. The major premise is a universal principle connecting the explanandum
with the explanans, the minor premise states the existence of the explanans, and the conclusion entailed is the explanandum. In a typical explanatory situation, the explanandum has been observed and the explanans is offered to explain why the explanandum occurred. Presupposed by the person offering the explanation is the universal principle connecting what is to be explained with what explains it. Presumably, when the recipient of the explanation is satisfied by the explanans, he or she has likewise presupposed the general principle connecting the explanandum with the explanans.

In terms of reason explanations, the major premise is a universal principle connecting what is to be explained (the action) with what is to explain it (the desires and beliefs). The minor premise consists in the existence of the desires and beliefs, and the conclusion entailed is the action. In a typical situation whereby an explanation is sought for an action, the existence of the desires and beliefs will be cited. It is only if this explanation fails to satisfy that the universal principle, presupposed by the provider but in this instance apparently not by the recipient of the explanation, will come to be mentioned explicitly.

It is thus central to Solomon’s (1974) argument that at least the very general universal principle that actions are caused by beliefs and desires be presupposed in the giving and receiving of reason explanations. For example, the success of an explanation of an action such as my opening the drawer in my bedroom in terms of my desire to fetch my socks, and my belief that they are in the drawer, relies on the presupposed and hence unmentioned principle that desires and beliefs determine behaviour. If such a principle were not presupposed, for example, if the persons to whom I was explaining my action by mentioning my desire and belief did not hold that desires and beliefs determine behaviour, then my explanation would more likely puzzle than enlighten them. It would be legitimate for such persons to wonder why the desires and beliefs had been mentioned in the explanation at all. Certainly knowing only that the desires, beliefs and behaviour all occurred is not enough to suggest, independently of the universal principle, that the desires and beliefs account for the behaviour because they caused it. Yet this is clearly what one intends to convey when citing either a motive or a belief, or both together, as an explanation of rational behaviour. As Solomon argues, “… the general laws supporting reason-explanations may be inappropriate for everyday
discourse because they are presupposed in context; from this it does not follow that no such general laws are applicable or necessary. On the contrary, their presupposition demonstrates that they are both applicable and necessary” (pp.427-428).

If Solomon (1974) is correct, then the success of a reason explanation requires, at least, the very general principle that ‘motives and beliefs together determine action’. In response to proposals of this sort, Davidson (1976) has argued that very general behavioural principles such as Solomon’s are conceptually central to the notions of belief, desire and behaviour and that they are therefore of more relevance to elucidating what is meant by rationality than to explaining why an action occurred.

Recall that, according to the practice of radical interpretation, content is attributed to the beliefs, desires and actions of another in accordance with what it would be rational for that other to think, desire and do. Furthermore, the attribution of content requires an holistic approach whereby beliefs, for instance, are attributed content in accordance with other beliefs, desires and behaviours. Similarly, behaviours are interpreted and attributed meaning in the light of other behaviours, desires and beliefs. Because rationality is assumed in psychological attributions, Davidson deems it unenlightening to suggest that reasons cause behaviours as if this were something beyond a methodological constraint guiding content attribution. Concerning a general behavioural law such as, “Anyone who is rational with respect to certain beliefs and desires at a certain moment will act on them”, Davidson writes that, “It does not seem likely that we can give a testable content to the required relativized concept of rationality that would allow us to test this law empirically. But if we cannot, the so-called law merely states part of what we mean, in this context, by saying an action is rational” (1976, p.267).

If indeed the content or meaning of a rational action derives from its conceptual ties to the desires and beliefs that cause it, then it is circular to appeal to those same desires and beliefs when seeking an explanation of the action’s occurrence. Because Davidson assumes that the contents of an individual’s attitudes and actions are logically intertwined, he can agree with Solomon that presupposed in a reason explanation is a general principle connecting actions with the desires and beliefs that cause them whilst refusing to acknowledge that this general principle has anything to add to a reason explanation that is not conveyed simply by stating that the beliefs, desires and
behaviour all occurred. Davidson’s view that very general behavioural principles are redundant in reason explanations therefore turns on his assumptions that, firstly, it is only as a result of interpretation and attribution that beliefs, desires and actions have content at all, and, consequently, that any general laws relating to attitudes and actions are conceptual givens.

The disparity between Solomon (1974) and Davidson (1976) with regard to the necessity or otherwise of behavioural principles for reason explanations has so far been considered only with reference to principles that are so general as to apply to all instances of behaviour. Davidson has rejected such principles as irrelevant to the explanation of actions in terms of reasons because of what he sees as their inherent circularity in that role. Solomon, for different reasons, recognises that principles of such great generality may be inadequate for the explanation of specific actions. He writes that “… this most general principle may yet be inadequate to explain specific actions, for what is required in a reason-explanation is not just the statement that motives and beliefs are involved – this simply says that a reason-explanation is possible – but a statement of the specific motives and beliefs involved. The general law involved must therefore deal with these specific reasons, not with reasons in general” (pp.424-425). In other words, the very general principle that ‘beliefs and desires together determine behaviour’ is so non-specific as to provide only a rather anaemic account of why someone behaved in a certain way. To remedy this, very general principles need to be made more specific by citation of the particular action to be explained, and the beliefs and desires that will explain it.

What happens, however, as soon as content is specified in behavioural laws is that they are opened to falsification through instances whereby an individual has the requisite desires and beliefs and yet fails to perform the action. For example, while it may tend to be the case that persons desiring their socks who believe their socks to be in their bedroom drawer will open the drawer in search of the socks, some individuals will have the desire and belief combination, and yet not open the drawer. Indeed, Davidson (1976) argues that, “Far more often than not people fail to perform any action at all to achieve a desired end, even though they believe or know the means are at hand” (p.263). Thus, it appears to Davidson, the closest one can come to a behavioural law is a tendency, “… we can say of someone who has a desire or end that he will tend to
behave in certain ways under specified circumstances” (pp.263-264). Tendencies, however, are rather loose versions of laws. Rather than universal principles connecting reasons and actions, Davidson’s notion of a behavioural law suggests only that, “If the agent had the appropriate belief and desire, it was at least possible for him to perform the action to be explained, and there was some (very low) probability that he would” (p.264). A question then arises concerning why behavioural laws, unlike those appearing in ideal explanations, are such that it is to be expected that many exceptions to them will be found.

The answer, unsurprisingly, is tied up in Davidson’s (1976) view that “… the laws implicit in reason explanations are simply the generalizations implied by attributions of dispositions. But then the ‘laws’ are peculiar to individuals, and even to individuals at particular moments” (p.265). Breaking this answer down into its component statements it can be seen that it is because laws are deemed to be artefacts of content attributions that they have very limited application, indeed, that they appear only to apply to one individual once. For example, in explaining why someone has performed an action such as opening a drawer, an interpreter ideally takes into account the whole network of the actor’s attitudes, other behaviours, and any other signs that could be indicative of what the action means, such as utterances. Taken as a whole, the actor’s psychological makeup is interpreted so as to reveal a rationally coherent set of attitudes and actions, to the best of the interpreter’s ability. However, as time passes and new evidence becomes available, for example in the form of a new utterance, prior interpretations will be discarded in favour of ones that are better suited to the evidence as a whole. As content attributions alter, so too will the explanatory behavioural generalisations. Therefore, behavioural generalisations are likely only to apply to individuals at particular moments.

From this it follows that any attempt to extend a behavioural law’s application, even to the same individual at a later date, is illegitimate, and it should therefore be unsurprising that exceptions to the law are readily found when it is applied to individuals for whom it is not intended. Consequently, it is because of their tentative and individualistic nature that more specific behavioural laws produce reason explanations that are less than ideal, and even low-grade, a situation that Davidson finds satisfactory given the anomalous
character of mental phenomena. Yet this acquiescence to low-grade explanations that rely on ‘one-off’ laws is, if Solomon (1974) has reasoned correctly, unacceptable.

Recall that, according to Solomon (1974), the practice of explanation requires that there be some principle connecting explanandum with explanans presupposed by both the giver and receiver of the explanation. Where no principle of this kind is presupposed, then simply mentioning that a person had certain desires and beliefs can be of no use in explaining why he or she acted accordingly. Indeed, the notion of ‘acting accordingly’ is not even intelligible if there is no strictly lawful connection between the desires, beliefs and behaviour. For example, a reason explanation that appeals to the non-strict law that ‘often, but not always, persons desirous of their socks who believe their socks to be within their drawer in their bedroom will open their bedroom drawer in search of the socks’ is equally applicable to persons with the requisite desires and beliefs who do not open the drawer as to those who do. The fact of having the desires and beliefs could equally well explain one person’s action and another’s inaction. As such, citation of the desire and belief combination is unenlightening as an explanation of why one party acted ‘accordingly’, while the other did not.

Arguments similar to these have drawn from Davidson (1963) the suggestion that the rough generalisations involved in reason explanations are not completely divorced from the kinds of strict laws one expects to find in explanations; indeed, they are deemed to “… provide evidence for the existence of a causal law covering the case at hand” (p.683). Nevertheless, it has been Davidson’s consistent claim that the laws covering the case at hand need not be formulated in the same vocabulary as the items to be explained. All that he claims is that “…if A causes B, there must be descriptions of A and B which show that A and B fall under a law” (1976, p.262). As has been spelt out, these descriptions need not, and indeed cannot, be mental descriptions if the law is to be strict (see Section 2.3.2). In other words, the relevant strict law will feature redescriptions of the beliefs, desires and actions in physical terms. Thus, while an applicable strict law connecting the explanandum and explanans is discoverable, it is a strict physical law, not a psychological one.

Despite Davidson’s concession to the indirect relevance of strict laws to reason explanations, concerns remain that in the absence of strict psychological laws, reason
explanations remain mysterious. For example, Dagfinn Føllesdal (1985) argues that, “If the only law that connects A and B is instantiated by A and B only under descriptions that differ radically from ‘A’ and ‘B’, then the assertion that A is a cause of B will be more the pronouncement of an oracle than it will be an explanation” (p.315). In other words, the possibility that when redescribed the events in question would fall under a strict physical law is of no relevance to someone seeking an explanation in terms of reasons. Antony (1989, 1995) argues along similar lines. She contends that, “… while it may not matter for the truth of a singular causal claim how the cause is described, it does matter for the adequacy of a causal explanation. Causation may be extensional; explanation surely isn’t” (Antony, 1995, p.430).

The problem with proposing a nomological account of explanation whilst maintaining that the law of the major premise and the assertion of the minor premise may be differentially characterised is that no inference can be derived from the premises’ combination. As argued by Julia Tanney (1995),

… Davidson’s move to go ‘extensional’ with causation, but remain ‘intensional’ with laws, robs us of the explanatory connection between the laws and the particular events that the laws putatively subsume. This is because the ‘appropriate’ event description will surely be one in which the implication goes through – namely, one in which the events referred to in the singular sentence of the minor premise are picked out by the same predicates as the event (types) referred to by laws in the major premise. But if the vocabulary in which the laws are couched (a physical vocabulary) is different from the vocabulary in which the events that stand in a causal relation are picked out (a mental vocabulary) and, as his psychophysical anomalism implies, if there are no bridge laws connecting the two domains, then it is difficult to see how laws described in a physical vocabulary would be able to do the requisite ‘subsuming’ of events described in a mental vocabulary (pp.117-118).

Arguments such as those of Føllesdal, Antony, Tanney and others14 reveal that the apparently successful application of the extension-intension divide in allowing mental events to be both causally efficacious and anomalous is of no use for demonstrating that

14 See also Hedman (1973) and Stoutland (1986).
rationalisations can be successful in the absence of psychological laws. This is because, for the purposes of explanation, how the cause and effect are described is of the utmost importance. Thus, the claim that reason explanations are justified because their constituents, differently described, reveal the relevance of strict laws, is an unsuccessful contender for replacing the view that reason explanations are justified because they instantiate strict psychological laws.\(^\text{15}\)

Interestingly, it seems possible that of late Davidson (1995a) has come to accept something like this point of view. While maintaining that, “Given the endless possibilities for redescribing events (or anything else) in non-equivalent terms, it is clear that there may be no clue to the character of an appropriate law in the concepts used on some occasion to characterise an event”, he concedes that, “What may be the case is that if a singular causal statement is to be explanatory in some desired sense, it must put its hearer in mind of at least the general nature of a relevant law” (p.265). Clearly this is achievable only when the relevant law uses concepts similar to those contained in the singular causal statement, not conceptually unrelated redescriptions.

It thus appears that if Davidson is to maintain that reason explanations are viable, he must also acknowledge that this is because there are strict laws, formulable in psychological terms, connecting the explanandum and explanans in rational action. Of course, this acknowledgement undermines part of what is meant by the claim that mental phenomena are anomalous because it admits that there are, or can be, strict deterministic laws by which mental events can be explained.

Perhaps even more importantly, it opens the door for speculation as to the character of psychological content and behavioural laws. For example, to Davidson’s argument that very general behavioural laws merely express what is meant by rationality it may be objected that whether or not actions are lawfully connected to the desires and beliefs that rationalise them is an empirical matter, one that can only be discovered through observation of how people do in fact act, given their desires and beliefs, and thus that it is not conceptually circular to find that certain desires and beliefs cause certain actions.

\(^{15}\) This conclusion is not, it must be noted, universally accepted. For example, in defence of Davidson, Barrett (1994) questions the necessity of psychological laws in reason explanations as follows, “... it is not even clear that we should accept Føllesdal’s claim that a law, hedged or strict, must relate a particular property of a cause and a particular property of an effect for reference to those properties to make significant explanatory contributions to an explanation of the effect” (p.84).
To Davidson’s thinking, this separation of observation of the relations between attitudes and actions from the content of those attitudes and actions would be a nonsensical one, because the content to be attributed is determined by practices of observation and interpretation. Nonetheless, it is a position with merit if it can be shown that the content of actions and attitudes is not reliant upon third person interpretation and attribution, and the possibility of psychological laws suggests just this. I shall elaborate.

According to radical interpretation, the content of, for example, a belief, is deemed to be the product of an observer’s interpretation of an agent’s utterances, environmental context, other attitudes and so forth. As such, content attribution is highly individual and varies with the capabilities, biases and so forth of the interpreter. Just as beliefs are deemed to be attributed content in this manner, so too are desires and actions. Because of the assumed holistic and normativistic character of an individual’s attitudes and actions, the content of an attitude or the meaning of an action is open to alteration according to what the interpreter deems to be the most rational overall interpretation possible. It is in this sense that the concepts of belief, desire and so forth are understood to be attributed with content in accordance with an evolving theory that aims for maximum rationality. Given the tentative nature of content attribution, it is exceedingly difficult to imagine how psychological laws, universal principles concerning contentful psychological items, could be developed. Naturally, if the constituents of a psychological law, such as the content of a belief, can be altered at any time, then the law itself is unsustainable. It thus follows that if psychological content is a product of radical interpretation, then psychological laws, understood as something more substantial than tendencies, are impossible to formulate. Therefore, the purported existence of psychological laws suggests that the content of attitudes, and the meaning of actions, are not dependent upon third person acts of interpretation. Furthermore, if attitudinal content is independent of interpretation, so too must be the laws that reveal the connections between attitudes and actions. Rather than being an artefact of human interpretative and linguistic processes, it may be hypothesised that both psychological laws and attitudinal content are facts of nature existing independently of third person interpretation. This suggestion will be taken up in Sections 3.4.2 and 3.5.
3.4 Mental properties as linguistic artefacts

Supervenient mental properties, or, alternatively, propositional attitudes, are understood by Davidson to be constituted by the mentalistic descriptions that we humans use to pick them out, “… the attribution of thought to others is a matter of matching the verbal and further behaviour of others to our own propositions or meaningful sentences” (1996, p.167). The content of a propositional attitude is supplied, Davidson (1989b, 1993a, 1993c, 1997) has argued, by the sentences uttered or written by interpreters. Thus, he has claimed, “whether an event can be called mental… depends entirely on how that event can be described” (1993a, p.8). Furthermore, the rational structure of attitudes and actions with each other is seen to derive from the constitutive rationality of the Unified Theory, with its inbuilt assumptions of holism and normativity. Thus, both the content and internal logical structure of attitudes and actions are deemed by Davidson to be constituted by the descriptive sentences that are employed to characterise them. This is the situation that I have characterised above as ‘constitutive description’ (Section 2.2), and in relation to which Stephen Sommerville (1980) recognises Davidson’s employment of ‘semantic ascent’, which is a term of Quine’s referring to, “… that practice, by no means unique to philosophy, in which the focus of an investigation shifts from the phenomena themselves to the terms in which the phenomena are described” (p.47). It is the aim of what follows to demonstrate that nothing is or could be constituted by being described, indeed, that only through having an existence that is independent of description can something come to be described, and thus that Davidson’s linguistic criterion for the discernment of mental from physical properties is unlicensed.

3.4.1 The logic of symbolism

As a linguistic act, description is a variety of conventional symbolisation. As such, it is subject to what Agnes Petocz (1999) refers to as ‘the logical constraints of symbolism’, of which she identifies four. The first of these, with the three implications that can be derived from it, she depicts as follows:

The most important logical constraint is that symbolisation (conventional or non-conventional), like any signification, is a relationship between three independently existing and independently characterisable terms: the signifier (in this case the symbol), the signified (in this case the symbolised), and the subject
(for whom the symbol stands for, or substitutes for, the symbolised). There are three noteworthy implications of this constraint: (i) firstly, the three-term relation cannot be presented as, or converted into, a two-term relation, by collapsing any two of the terms, particularly, as does happen, by collapsing the signifier and the signified; (ii) secondly, because meaning is a relation (X ‘means’ Y to A – the symbol ‘means’ the symbolised to the person), meaning requires each of the three terms, and cannot therefore be a property or quality of any one of them; (iii) thirdly, the symbolised must exist independently of its role in signification, i.e., it must have intrinsic properties which are neither created nor changed by the signifier, or by the fact that it is signified (p.242).

Consider an example of written language (all claims about which can be generalised to spoken language). The terms of the system are as follows:

a) the marks on the page or syntax, for example, the lines and curves that make up the word ‘red’ (the symbol);
b) this colour (the symbolised): ; and
c) the person for whom the symbol ‘red’ is used to stand for this colour (i.e., an English speaker; the subject).

How, then, are Petocz’s implications realised?

i) All three terms (the symbol, symbolised and subject) are necessary for a successful symbol system. For example, although a non-English speaker could recognise both the symbol (‘red’) and the colour symbolised ( ) just as well as an English speaker, knowledge of the relationship between them, that the word ‘red’ is used in English to stand for this colour ( ), would be lacking. Without the third term, the person for whom ‘red’ stands for ‘ ’, the symbolisation cannot take place. Similarly, the absence of either the symbol or the thing symbolised renders the attempt at symbolisation futile. Thus, all three terms are essential for symbolisation to take place.

ii) As can be seen when the non-English speaker does not know which colour the word ‘red’ is used to symbolise, the meaning of a linguistic symbol is not inherent in it.

iii) It can be seen readily that this colour retains all of its properties when it is symbolised ‘red’ in English, ‘rød’ in Danish, or however else it is symbolised in other languages. It is in no sense created or altered by the act of symbolisation,
rather, its various labels were developed following its discovery. In other words, the thing symbolised is possessed of positive characteristics independently of being symbolised.

Davidson’s view that ‘events are mental only as described’, which is central to his use of constitutive description, violates all three implications of symbolism being a ternary relation. By neglecting to acknowledge that mental properties exist independently of being described, Davidson:

i) collapses the symbol (mentalistic words) and the symbolised (mental properties of states and events) into one term;
ii) treats the meaning of mentalistic words and sentences as inherent properties of them; and
iii) suggests that the things symbolised (i.e., the mental properties) are created by the act of symbolisation (description).

These violations of the logic of symbolism will be considered in turn:

i) the collapsing of the symbols and the symbolised

At first glance it might appear that Davidson’s event descriptions comply with the necessity for three independent terms by recognising the symbols (mentalistic descriptions), the things symbolised (mental states and events) and the persons for whom the symbols indicate the symbolised (the interpreters). Nevertheless, this is a misleading picture. While Davidson takes pains to stress that events, like objects, are ontological items that exist indifferently to human characterisations of them, mental properties are deemed to be conceptual in character, and thus dependent for their being on human conceptualisations and, more specifically, human descriptive practices (recall that Davidson draws no distinction between concepts, properties and predicates). By denying a status to mental properties that is independent of the words by which they are indicated, Davidson has collapsed together the symbols used for mental attributions (the mental vocabulary) and what is being symbolised (mental properties of otherwise independently existing states and events).
ii) the treatment of meaning as inherent in mentalistic words

The assumed dependence of mental properties on the mentalistic vocabulary suggests that the words and sentences that constitute the latter bring with them their meanings. For example, in exploring what entities might be adequate for the assigning of content to another’s attitudes in accordance with the inbuilt standards of rationality of the Unified Theory, Davidson suggests that, “What is required is some potentially infinite supply of entities with a pattern or structure complex enough to provide a model for the attitudes” (1995b, p.11). As has been seen, the entities that are deemed to bring the desired content and rational structure to the attitudes and actions of others are interpreter’s utterances (see Section 2.2.2). Utterances, then, are understood by Davidson to be the appropriate instruments of content attribution. However, by allocating them this role, Davidson treats mentalistic sentences as if they contained their own meaning independently of being used by persons to indicate mental properties. Furthermore, by appropriating them as models for the attitudes, he assumes these sentences to have a complex rational structure. Accordingly, by arguing that mentalistic sentences are the appropriate tools for providing structure and content to actions and attitudes, Davidson has implied that those sentences contain content and structure as characteristics inherent in them.

iii) the suggestion that the things symbolised (i.e., the mental properties) are created by the act of description

By claiming that ‘events are mental only as described’ and that ‘the mental is not an ontological but a conceptual category’, Davidson proposes that mental states and events do not have permanent mental characteristics that exhibit stability independently of the vicissitudes of descriptive practices. Mental attributions, we are told, evolve, and, furthermore, there may be no way, even in principle, of deciding between alternative interpretations and attributions that are incompatible with each other. All of this suggests that there is no fact of the matter concerning what mental characteristics any ‘mental’ state or event possesses. Mental properties are accordingly understood to gain their being through being described, and thus are deemed to be created by the act of symbolisation. Consequently, the logical necessity that the phenomena being symbolised have their own characteristics that are neither created nor changed through being symbolised is violated.
3.4.2 Implications of rejecting ‘constitutive description’

The situation whereby the characteristics of objects and events are deemed to be determined by the application of a vocabulary to them suggests that vocabularies bring with them their own characteristics that can subsequently be imposed upon the objects and events to which they are applied. This notion, however, has been shown to be problematic because it suggests that what are in fact arbitrary symbols have inherent meaning and that mental properties fail to have an existence independently of the psychological attributions that ‘pick them out’. Furthermore, the denial of the independent existence of mental properties, taken with the view that mentalistic words are inherently meaningful, results in the collapsing of the words and the properties together, with the latter being constituted by the former. Because it violates the logic of symbolism in these ways, the doctrine of constitutive description is untenable, and the theory that mental properties are conceptual and linguistic entities arising out of the supervenience judgements of interpreters must be rejected.

In place of viewing mental properties as being constituted by mentalistic descriptions, two courses of action are available. The first of these involves accepting that belief in belief and the other mentalistic characteristics of human life must be erroneous because illusory, there being no such characteristics in existence. Of course, adoption of this course brings with it an end to the search for a theory of mental causation, because only things that exist can be causally efficacious. Disappointing as such an ending would be, it could not be avoided if there were indeed nothing to which our mentalistic words referred, no referents which were named by psychological terms. Fortunately, however, the attempt to adopt such a course of action reveals its futility. One can not ‘accept that belief in belief is erroneous’ without acknowledging simultaneously that there exists the possibility of believing falsely, without acknowledging, that is to say, that belief in the existence of belief is not erroneous. The self-contradictory nature of this first suggested alternative to the view that mental properties are mere descriptive artefacts therefore reveals it to be untenable.

The second possible course of action involves recognising that, rather than possess inherent psychological meaning, the mental vocabulary is distinguishable from the physical vocabulary because it indicates phenomena that are so distinguishable, that is, that possess distinctively psychological characteristics. It follows from this option that,
rather than an object or event’s membership in a type such as ‘mental’ or ‘physical’ being determined by the vocabulary by which it is picked out, the inclusion or exclusion of words in or from a particular vocabulary is determined by the characters of the things to which the words refer. If they refer to mental phenomena such as beliefs, then they belong to the mental vocabulary. If they refer to non-mental phenomena, then they belong to a vocabulary other than the mental one. Not only because the first suggested course of action involves itself in self-contradiction, but also because it is observable that people do have beliefs and other propositional attitudes, it is clear that it is this second course of action that must be adopted, that is, that it must be accepted that, whether ever described or not, events of a cognitive character occur and therefore that, in the inventory of real-world existents, mental characteristics take their place.

The implications of recognising that mentalistic language reflects rather than imposes itself upon the world are of direct relevance to Davidson’s proposed solution to the paradox of mental causation, particularly as concerns mental realism, psychological externalism, psychological rationality, and the non-linguistic character of laws. Perhaps the first and most obvious of the implications is the necessity for rejecting Davidson’s anti-realist conception of mental phenomena, whereby an event is mental only if it is so described or attributed through a supervenience judgement.

Clearly, because the act of making a judgement concerning the mental happenings of another is itself a mental act, consistency requires that this could not occur without observation of the observer and judgement of the judger. In this we see the beginnings of an infinite regress, one in which no observation or judgement concerning the mental acts of others could ever be made, because such observations and judgements, themselves being mental acts, could not occur until they, in their turn, had been observed and judged, and these latter observations and judgements could not occur until they had been observed and judged, and so on in a chain of observations and judgements without end. It thus follows that if observation and judgement, together with the acts of description that follow from them, are necessary conditions for the existence of mental acts, then no mental acts could ever occur.

In this we see that the problem with the assumption that events are ‘mental only as described’ is that it confuses the process by which we come to know what others are thinking and doing (this process involving observation and inference), with the thoughts
and deeds themselves. From this false step it is concluded that if there is no-one in a position to judge and describe what another is thinking and doing, then that other is unable to think or do anything. But, of course, one can only observe and thereby become acquainted with the thoughts and deeds of others if such thoughts and deeds are occurring already, independently of the observation that is turned upon them.

Acceptance of the counter position to Davidson’s, that if an event is mental in character, then it is so regardless of how it is evaluated and described, involves recognising that events have properties, and hence are universal or belong to ‘types’, independently of human concerns. That events have properties, and therefore universality, is considered in Section 3.5 below. Suffice for the present to remark that although events may well be amenable to a variety of descriptions as Davidson claims, their amenability reflects their characters that exist prior to, and independently of, such descriptions.

The second implication concerns psychological externalism. Davidson considers his theory to be an externalist one because, according to it, in attributing content to the thoughts of others interpreters make use of external objects that are accessible to interpreter and interpretee alike (as indicated under the heading ‘triangulation’ in Section 2.2.3). Nevertheless, despite his stated commitment to psychological externalism, Davidson simultaneously maintains that mental states are ‘inner’, as is required for an identity theory between mental and physical states and events. Accordingly, he argues that, “… mental states are characterized in part by their relations to events and objects outside of the person, but this does not show that mental states are states of anything more than the person or that they are not identical with physical states” (1986, p.167), and again, “It should be clear that it doesn’t follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren’t in the head” (1987a, p.451).

However, it is doubtful that this dual sense of identity, whereby mental states are conceived to be ‘inner’ because they are bodily states, and ‘external’ because they are about external objects, can be maintained. As argued by Andrew Woodfield (1982), “No de re mental state about an object that is external to the person’s brain can possibly be identical with a state of that brain, since no brain state presupposes the existence of an external object. Any state which did incorporate an environmental object would not
be a state of the brain, but would be rather a state of the brain-environment complex” (p.viii). Convincing as this argument is, Davidson (1987a) has a reply that should not by now be wholly unexpected. He concedes that it is not actually the brain states themselves that presuppose external objects, rather it is interpreters’ descriptions that do, “Individual states and events don’t conceptually presuppose anything in themselves; some of their descriptions may, however” (p.452). But, as has just been established, nothing exists by virtue of being described, and it thus follows that no amount of conceptualisation by an interpreter could ever furnish an internal bodily state with external mental content. Accordingly, because Davidson’s externalism relies upon the faulty notion of interpreter-based attributions of content to what are otherwise internal bodily states, it is likewise faulty. The key to resolving this conundrum, it will be argued in Part II of this thesis, requires embracing a consistently externalist account of cognition, whereby the theory that persons possess mental states is replaced by the theory that persons partake in mental relations with the world around them.

The third implication of recognising that mentalistic language reflects rather than imposes upon the world concerns the source and ‘necessity’ of psychological rationality. It has been noted that Davidson understands psychological rationality to derive from the norms and standards dictated by the Unified Theory with its basic assumptions of normativity and holism. Meanings, actions and attitudes are claimed to be attributed content so as to construe individuals as thinking and acting with as much consistency as an interpreter’s charity allows. But, of course, the Unified Theory, like any theory, is comprised of words and sentences that are not inherently meaningful and of which it would be nonsensical to claim have a rational structure. Rational structures may be reflected by words, for example as occurs with syllogisms in their written and spoken forms, but they cannot be derived from them. This casts doubt on one of Davidson’s fundamental assumptions, that the network of propositional attitudes must be rationally constituted. What had appeared to be a matter of definition, a requirement for the proper usage of mentalistic words, now turns out to be something quite different.

Whence, then, human rationality? This question has been answered to my satisfaction by John Maze (1991), who asserts that, “The court of appeal for validity is the structure of the world itself” (p.178). What Maze means by this assertion is that it is in the organisation of the mind-independent world that we are to find the rules of logic, the
laws, for example, of non-contradiction and entailment. For an example of the latter, consider the validity of the inference that ‘Socrates is mortal’ from the premises ‘All men are mortal’ and ‘Socrates is a man’. That the conclusion that ‘Socrates is mortal’ follows from these premises has naught to do with anyone’s inbuilt psychological rationality, it is a fact about the world and the mortality of some of its inhabitants, men. It follows from the fact that logic is a feature of the world, that logical thinking is simply an instance of knowing something about the world, that, to continue with the Socrates example, the following is known: ‘if all $p$ are $q$, and $x$ is $p$, $x$ is $q$’. As Maze suggests, what is usually termed ‘rational thought’ is “… simply the direct apprehension of a complex fact in the world, the fact that if certain pairs of properties are joined together in a particular way then there must be, and is, a further conjunction of them even if we did not pay attention to it immediately” (pp.178-179).

In recognising that rationality is not an inherent and inevitable feature of thought, that, although a thinker might apprehend complex facts such as those involving entailment between situations, he or she might not, we come to see that it is an open question whether persons can hold contradictory beliefs and act in ways other than those that best serve their desires. Furthermore, even if, as a matter of contingent fact, most people most of the time do think and act rationally, the source of their thinking and acting so must be sought in conditions that lie beyond the character of thought and action itself, and is most likely to be found to reside in the survival benefits that thinking and acting in compliance with the organisation of nature offers. Accordingly, psychological rationality, where it obtains, is best understood not as an inherent feature of thought and action, nor as a consequence of the prescribed methodology of psychological attributions, rather, as a demonstration of one’s sensitivity to the logical structure of the world.

The fourth implication concerns Davidson’s claim that ‘laws are linguistic’. It was seen in Section 2.3 that Davidson understands causal explanations, of which causal laws are a species, to be formulated in terms of events ‘as described’ in one way or another. It was this that separated explanations from singular causal statements, the latter being indifferent to the type of description used. Physical descriptions of causal relations, it had been claimed, are amenable to refinement into strict laws. Non-physical descriptions, and in particular mental descriptions, are, on the other hand, irretrievably
imprecise. However, if the argument above is correct, then only the names of laws, and not the laws themselves, are linguistic. For example, the words ‘law of gravitation’ name a non-linguistic regularity in the world whereby, “the attractive force of bodies varies directly as their masses and inversely as the square of the distance between them” (Oxford Illustrated Dictionary, 1976, p.737). To repeat, the name of the law, not the law itself, is linguistic. It thus follows that events and laws cannot be placed on either side of the ontological-conceptual divide as Davidson had hoped. If indeed there are no psychological laws, then this is a fact of nature, not, as Davidson has argued, a consequence of the normative character of psychological attributions. If there are psychological laws, then these might, in the future, be discovered and even named. However, at no subsequent point would such laws themselves take on a linguistic character.

The proposal that mental properties and the laws that are formulated in terms of them are conceptual and linguistic is thus unable to withstand close examination. Rather than furnish events with mentalistic characteristics, it is concluded that psychological descriptions are used by persons in order to reflect discernible characters of events that do, and must, exist independently of their being so discerned and described. Similarly, whilst the naming and describing of lawful regularities is linguistic in the sense that both are forms of communication that utilise written and spoken symbols, the regularities themselves are discoverable facts of nature that need not be named or described in order to occur.

3.5 The particularity and universality of events

The crucial assumption upon which the theory of Anomalous Monism rests is that while events are particulars or individuals, properties and the laws that pertain to them are universal, conceptual, and linguistic. It has already been argued that nothing can be constituted by being described and therefore that all properties, together with their relations (whether lawful or otherwise), must exist independently of description if they are to be described. Having put aside as false the notions that ‘events are mental only as described’ and that ‘laws are linguistic’, it is time now to consider the viability of the view that events are ontological particulars in the sense of being devoid of universal features or properties.
It is not at all difficult to understand why Davidson would embrace the notion that events, but not properties of events, are ontological items. The separation of events and properties into ontological and conceptual categories respectively appears to provide the flexibility required for mental events to be both causally efficacious and free from strict laws. This proposed solution to the paradox of mental causation serves only, however, to breed new and more perplexing questions, namely: if events are not possessed of properties, how do they gain their identities, and, relatedly, whence their causal efficacy?

Despite the centrality of events to Davidson’s theory, and the existence of several of his articles concerning them, there is very little material available by which one can discover of what, exactly, Davidson deems events to be comprised. As was detailed in Sections 2.1.1 and 2.1.2, Davidson takes events to be both fundamental ontological items and particulars, but, beyond this, his characterisation is carried out in terms of how best to individuate events (that is, how to tell them together and apart). This practice should, one would suppose, throw light on what characteristics of events contribute to their identity, but it will be noticed that in his efforts to provide individuating conditions of events Davidson looks not to their properties, but to their relations, formerly their causal relations and more recently their spatio-temporal relations, “… events, like physical objects, are identical if they occupy the same places at the same times” (1985a, p.175). Accordingly, Davidson’s readers are left knowing only that, in his view, events occur in time and space and that they owe none of their identity to any of the properties that are imposed upon them. This neglect of properties in the process of event identification is, of course, deliberate. If indeed events could be identified by their properties, then they would not be particular, ontologically speaking. If properties belonged to events themselves, and were not an artefact of human interests, then it would have to be conceded that, in addition to their particularity in the locational sense, events also possess ontological universality in that they have characteristics in common with other events. Furthermore, if events were comprised of properties, then the question of the relevance or irrelevance of these to their causal relations would become pertinent, and the criticisms canvassed in Section 3.2 concerning mental property, and hence mental event, epiphenomenalism, would come into force. Consequently, Davidson must reject the view that events are comprised of properties (in an ontological sense) if his solution to the paradox is to be successful.
Yet no event, I contend, is particular in the sense of being without universal features. This is not to deny the relevance of events’ particularity in a locational sense, each event occurs at a time and place, but to assert that each event is, ontologically speaking, also universal in that it possesses types of properties or characteristics. For example, the birth of Jane Austen, as was claimed earlier (Section 2.1.2), is a particular event in that it is dated, placed, and unrepeatable. It is also universal in that it is a *type* of event, the birth of a human child, and this latter fact owes nothing to human conceptualisation or description. It is because they share relevant features that all births can be collectively described as such, but the occurrence or non-occurrence of descriptions is immaterial to these events being the universal type (births) that they are.

Similarly, all mental events share necessary and sufficient features that gain them admittance into the class ‘mental event’. For example, as Davidson himself has often remarked, each of the propositional attitudes has a belief component. Again, this is not (as Davidson would suppose) simply a way of speaking; it is a fact concerning organisms and their cognitive relations to the world about them. Mental events form a class because they have certain characteristics in common.

Despite this, the particularity of events that Davidson’s theory requires involves them having no universal features, ontologically speaking, that are necessary and sufficient to gain them membership in a type. But if events were truly devoid of universal features in this way, if they were ‘ontologically neutral’, then they would consist of nothing at all. Such a conception of events is highly problematic, firstly because it is incompatible with the notion of token identity required by Anomalous Monism, and secondly because it is unsustainable. The token identity problem was recognised early on by Noren (1979), who is troubled by the “… basic assumption that a given event can be either mental or physical depending on whether it falls under a mental or physical description” (pp.70-71), because he is uncertain as to what could constitute ‘a given event’ under these circumstances. He expresses his concerns as follows,

Let us assume it makes sense to say that intentional-mental discourse, talk of persons as wanting, willing, intending, believing, sensing, and the like, and physical discourse, talk of persons in the terms and predicates of the natural sciences, comprise two independent languages, concept systems, or theoretical
frameworks for describing persons and their behavior. On our theory, each concept system proceeds on the basis of its own special idioms, categories, and nomologies. No causal laws can be framed across the two. They are thus autonomous frameworks which, by virtue of their conceptual differences, generate descriptions of persons having different logics, descriptions which thus exclude each other.

But now, if this is the case, what reason is there to expect that each system will organize or structure its ontological units in the same fashion? That is, each system appears to differ dramatically as to how persons are to be ‘broken up’ into so many ontic units. So one wonders what sense can be made of the notion that an event *simpliciter* can fall under both mental and physical descriptions (p.71).

It thus appears that the differing ontic categories of each framework generate descriptions of persons that parse them into vastly different events, things, and their relations. What is a distinct, separable event to one framework may not be on the other. And a description truly applied to a person on one may not, on the other, correlate with any appropriate description at all (p.73).

Noren’s (1979) concern here is that it is impossible to isolate an event as a single unit if indeed events are ontologically neutral (i.e., devoid of discriminating properties). In terms of the events that psychologists are interested in, Noren argues that, “… there may be no ‘neutral something’ between the picture of a person as an intentional agent and our concept of him as a complex of microentities” (p.72). In other words, there may be no neutral events to which various descriptions can be applied. Consequently, beyond the proposed conceptual and linguistic criteria, there appears to be nothing available by which events can be isolated into units and, furthermore, nothing to guarantee any coincidence of boundaries for the individual mental and physical events that Anomalous Monism asserts to be identical, yet conceptualises according to incongruous theoretical criteria.16

---

16 See Antony (1989), Pradhan (1999), Seager (1992) and Suzman (1980) for additional arguments establishing the impossibility of the token identity of mental and physical events under Anomalous Monism.
That to deny events’ ontological universality is to deny their existence is further confirmed when it is recognised that Davidson’s claim that events are best individuated by their spatio-temporal relations is unsatisfactory as an alternative source of identity for the supposedly ontologically neutral events, because nothing is constituted by the relations it enters into. The claim that some relations are constitutive of their terms has been forwarded by George Van Sant (1959), but, as examination of his arguments reveals, this claim is based upon a confusion between what something is, and the ease with which it can be distinguished from other things. For instance, having noted that according to “… standard practice in logic, a relation is designated by the symbolic notion aRb, where a is the referent, R is the relation, and b is the relatum” (p.27), Van Sant uses the paternity relation to illustrate his notion of constitutive relations. Of this relation he writes, “It is constitutive because that there is such an entity b to which a bears relation R (fatherhood) says something about a. In ordinary usage one says that a is the father of b. One and only one man can be a, once the relation R is defined and a b specified. Obviously b could have more than one value. What is actually the case is that a constitutive relation such as ‘a is the father of b’ is a definite description. The relation and the relatum constitute a definite description by uniquely determining the referent” (p.28).

While it is certainly the case that only one person can be the father of any particular individual, and that, therefore, the relation and the relatum together limit the potential candidates for the referent to one, this does not establish that the referent is constituted by the relation and the relatum. The referent’s unique relation to the relatum does, in this instance, function more effectively to distinguish him from others than do the bits and pieces that together constitute his body, but the constitution question is not one concerning the ease with which he can be distinguished from those around him, but of what he is composed. He is composed, like all people, of flesh and bone, and this fact is unaffected by the relations he enters into, including paternity relations. Therefore, the fact that it is at times easier to distinguish an object or event from other objects and events by its unique relations than by its composing properties should not be misunderstood to imply that such relations can take on the role of the properties,

17 For an independent argument tending to a similar conclusion see Heckmann (1992), who rejects Davidson’s use of spatio-temporal zones as the individuating features of events because, for any two individual zones x and y (where x ≠ y), there exists a zone z, which is the union of x and y, and which undermines the individuating capacities of the original two zones.
rendering the latter unnecessary as the building blocks from which objects and events are composed. Yet this is what it seems Davidson must do if he is to maintain that events have existence, because he denies the existence of the mind-independent properties of which events are more typically understood to be comprised.

Understanding, then, that an event cannot be constituted by its relations, and keeping in mind that he denies the mind-independence of properties, it becomes clear that unless Davidson can propose some alternative ‘event-stuff” out of which events are composed, the causal relations that Anomalous Monism has been developed to explain could not occur because the supposed events involved in them have been whittled away to nothing at all. The charge of epiphenomenalism is accordingly one that is inappropriately directed at Anomalous Monism because it takes as its starting point the view that Anomalous Monism is a viable theory of events, which it is not. Criticism of Anomalous Monism need not proceed so far as to consider whether the ‘events’ it is based upon are causally efficacious, nor whether the ‘properties’ it ascribes are or are not of causal relevance, because the theory undermines the reality of each. In fact, queries of these kinds tend rather to obscure than to expose the underlying problems with Anomalous Monism, because they presuppose the reality of events and properties in a way that is incompatible with the theory. The real culprit within the theory of Anomalous Monism, it turns out, is the assumption that events and properties can be separated into disparate ontological groupings, an assumption that ultimately undermines Davidson’s attempt to resolve the paradox of mental causation.

There is, of course, an alternative to viewing events as propertyless particulars, and this is to recognise that events are particular in a locational sense, but universal in the sense that they are composed of mind-independent properties, that is, properties that are on an equal ontological footing with objects and events. Recognition of events’ universality brings with it the dissolution of the ontological-conceptual distinction that has appeared to Davidson to be the key to resolving the paradox of mental causation, but it re-establishes a realist conception of both properties and events, one from which the business of searching for an alternative theory of mental causation may be begun.

In summary, by assuming that events form an ontological category but that ‘the mental’ constitutes a conceptual category, Davidson has understood ‘mental events’ to be a
marriage of nature and art. Yet this marriage, it is clear, could not be consummated because neither party has any real existence. Understood as linguistic artefacts, mental properties are unsuited to being genuine characteristics inhering in events, and events as ontological particulars are, in any case, un receptive to any such advances. Ultimately, then, Davidson’s reliance on the theory of Anomalous Monism as a means of solving the paradox of mental causation is untenable because it undermines the reality of the very events that were thought to be causally efficacious.

3.6 Concluding remarks concerning Anomalous Monism
The above considerations (Sections 3.1, 3.2 & 3.3) of criticisms that take Anomalous Monism combined with weak supervenience to be a viable theory of the nature of mentality has led to a number of conclusions. Firstly, Davidson’s version of mental supervenience avoids internal inconsistency. The use of ‘possible worlds’ is unnecessary as a device for revealing that indiscernible subvenient physical properties can determine discernible supervenient mental properties because this is consistent with supervenience as it was originally formulated.

Secondly, attempts to demonstrate that Anomalous Monism entails mental \textit{event} epiphenomenalism have been thus far unsuccessful because if, as is assumed in Anomalous Monism, there exists an ontological-conceptual distinction between events and properties, then Davidson’s approach to mental causation requires only to show that some causally efficacious events have mental \textit{descriptions}, not that those same events have observer-independent mental properties. As Neil Campbell (1997, 1998) argues, standard objections to Anomalous Monism have typically assumed a realist position concerning properties whereby properties are not only observer-independent, but also the entities by virtue of which events gain their causal efficacy. While this stance has much in its favour, and is perhaps the one that Davidson should adopt, it is nevertheless at odds with the position that he has adopted. Consequently, critiques of Anomalous Monism that turn a blind eye to its anti-realist stance towards properties have flown wide of the mark, arguing against the difficulties that Anomalous Monism would have encountered had it taken a realist stance towards properties, without first demonstrating the necessity of changing in this way.
Thirdly, whereas Davidson has typically understood his denial of strict laws in the domain of psychology to entail only that psychological explanations are necessarily low-grade, it has been demonstrated that no explanations whatsoever can be formulated without laws. This is because explanation in general involves the citation of an explanans which is only of explanatory use if its lawful connection with the explanandum exists and is known. Where no such connection exists, no explanation by citation of the explanans is possible. Consequently, either psychological explanation is impossible, or there do exist psychological laws.

In summary, what has emerged from Sections 3.1, 3.2 and 3.3 is that if the ontological-conceptual distinction between events and their property descriptions, as well as between causal relations and causal explanations, remains unchallenged, then while it can be shown that Anomalous Monism cannot provide a basis for causal explanation, its apparent status as a viable theory of mental causation remains intact. In other words, Davidson’s critics who accept Anomalous Monism as a viable theory of mentality have been unable, thus far, to argue effectively that it is incompatible with mental causation because they have assumed, but failed to demonstrate, that in order for a mental event to be efficacious, it must possess mental properties that are causally relevant. Understandably, this latter view, that if an event is mental, it has mental properties, and that if a mental event is causally efficacious, it is so because of its mental properties, has been deemed too obvious by many critics to be argued for, or even, for the most part, to be mentioned. Nevertheless, it is only through argumentation of this sort that Davidson’s theory can be refuted properly.

The criticisms of Sections 3.4 and 3.5 were aimed at demonstrating that, firstly, if there are such things as mental properties then these, like all properties, are characteristics that exist independently of how they are conceptualised or described, and, secondly, that all events have both particularity and universality because they are simultaneously uniquely located and characterised by properties belonging to classes or types. From these observations it is concluded that the proposed ontological-conceptual distinction between events and properties that underlies Anomalous Monism is unsustainable, and therefore that the theory itself must be rejected. Naturally, the rejection of the theory of Anomalous Monism means that Davidson’s solution to the paradox of mental causation
must be rejected likewise, and a new solution sought. This is the task of Part II of this thesis.
PART II: A DIRECT REALIST APPROACH TO MENTAL CAUSATION

In Part II of this thesis, I propose to introduce and review existing Direct Realist contributions to the area of mental causation (Section 4), and then, where necessary, to criticise and improve upon these (Section 5). As will be discovered, the theory of Direct Realism cannot rightly be claimed to belong to any one person alone, nor to be complete within itself. Rather, it is a theory that provides a firm basis from which one may approach questions such as the one to be considered in Section 5: can cognitive relational situations be causal? Prior to undertaking that consideration, however, it is necessary to explore realism itself, and to discover what, thus far, has been suggested in terms of a Direct Realist account of mental causation. This is the aim of Section 4.

Section 4: Introduction to Direct Realism

4.1 Realism

The theory of Direct Realism is part of the broader philosophical position known simply as ‘realism’ that was taught by the Scottish born Australian philosopher John Anderson, who in his turn was influenced by earlier realists such as Samuel Alexander. Realism is introduced nicely by the following, “Realism holds that things known may continue to exist unaltered when they are not known, or that things may pass in and out of the cognitive relation without prejudice to their reality, or that the existence of a thing is not correlated with or dependent upon the fact that anybody experiences it, perceives it, conceives it, or is in any way aware of it” (Montague cited in Anderson, 1927a, p.27).

The basic premise of Anderson’s realism is that reality consists in what exists in space and time, and that such existence is mind-independent. Accordingly, realism asserts that the objects of knowledge are real world situations that exist prior to, and independently of, being known. Furthermore, realism denies that there are different realms of existence, as well as the associated claim that some realms exist only relatively to others, either below or above them, as subservient or supervenient; these facts being reflected in the logical conditions of discourse. He writes that, “It is worth noting that all theories of higher and lower realities are stated in terms of the common reality we all know – and, indeed, can be stated in no other way” (Anderson, 1935, p.90).
In contrast to the position that there do exist different realms of being, realism maintains that there is just one way of being, existence in time and space, and, accordingly, neither degrees of reality nor truth, “Rationalistic theories of all sorts are distinguished from empiricism by the contention that there are different kinds or degrees of truth and reality. The distinguishing-mark of empiricism as a philosophy is that it denies this, that it maintains that there is only one way of being” (Anderson, 1927b, p.3). Furthermore, from the fact that all things exist together in space and time it follows that realism implies determinism, “… that anything in “nature”, as spatio-temporal, consists of certain “ways of working”, affecting and affected by the “ways of working” round about it” (Anderson, 1939, p.260), and therefore that rather than mental causation being in any way paradoxical, it is the notion that mental events might be excluded from causal interactions that is impossible to maintain.

Anderson wrote about a wide range of issues, a selection of which has been gathered and published in *Studies in Empirical Philosophy* (1962). Of the many important issues treated in this and other publications, those of the most direct relevance to the present discussion concern the characterisation of causation and cognition as relational situations occurring in time and space. These, of course, form part of the subject matter of Anderson’s general theory of relational situations, a critical feature of which concerns the acknowledgement that all terms that enter into relational situations require independent existences, both from each other and from their being so related.

In what follows, Anderson’s conceptions of causation and cognition will be reviewed in conjunction with the advances that have been developed by some of his students, as well as by some of their students. Prior to considering these topics, however, it will be well to introduce some of the conventional terminology pertaining to relational situations, as well as to outline Anderson’s argument for the independence of terms in relational situations. For the purposes of explaining the terminology of relational situations, Bertrand Russell (1920, 1937) is invaluable, and the following definitions are, for the most part, to be found in these two works.

18 As noted by Baker (1986), “Anderson oscillated between realism and empiricism (he regarded them as virtually the same position) as the best description of his general position” (p.1). This accounts for the presence of both terms in the present quotations.

19 For a more thorough account of Anderson’s philosophy, see Baker (1986).
4.2 Relational situations

The relational situations that are the simplest to comprehend involve two terms, and are known as binary. Binary relational situations are conventionally symbolised as follows, aRb, which in plain English reads ‘a stands in relation R to b’. For example, if Elizabeth Bennet is denoted a, Mr Collins denoted b, and the relation ‘being a cousin of’ R, then aRb reads ‘Elizabeth Bennet is a cousin of Mr Collins’. The referent in this situation, the term that the relation proceeds from, is Elizabeth. The relatum in this situation, the term that the relation proceeds to, is Mr Collins. The sense of the relational situation is given by its direction, from Elizabeth to Mr Collins. There exists also the relational situation that has the opposite sense to the original situation, that is, the one that proceeds from Mr Collins to Elizabeth (bRa). When this second relationship is implied by the first, the relation involved is known as the converse relation. In this particular instance, both the original and the converse relations are of the same type, ‘being a cousin of’.

Consider, though, another type of relation, ‘being the wife of’. Mrs Bennet, a, is the wife of Mr Bennet, b, and this situation may be symbolised aRb. The converse of this relation involves the introduction of a new type of relation, namely, ‘being the husband of’, R’, and the second situation is symbolised, bR'a. Note the difference that arises between the first situation (where Elizabeth is a cousin of Mr Collins) and the second (where Mrs Bennet is the wife of Mr Bennet) when their respective converse relations are considered. Both converse relations give the opposing senses of the original relations (the new relational situations proceed from the original b terms to the original a terms), but whereas in the first situation the type of relation remains constant, in the second situation it alters. This difference is what is referred to by saying that the first relational situation is symmetrical, whilst the second relational situation is asymmetrical. While symmetrical relational situations are compatible with their converses in that both senses involve the same type of relation, asymmetrical relational situations are incompatible with their converses in that they necessarily involve differing types of relations. In addition to symmetrical and asymmetrical relational situations there are non-symmetrical relational situations, these being those where the converse relations might, or then again might not, be the same kind of relation as was
Finally, while it is convenient to consider specific instances of relational situations such as Mrs Bennet’s being the wife of Mr Bennet, and thereby to specify the referent and relatum, it is necessary to acknowledge that any given relation, such as ‘being the wife of’, involves a class of possible referents (all wives) and a class of possible relata (all husbands). These are known respectively as the domain and the converse domain of a relation, and they together comprise its field.

It was mentioned above that Anderson’s characterisation of relational situations as requiring independent terms is of critical relevance to the present discussion. By ‘relational situations’ I mean to indicate situations whereby two or more terms stand in a relation to each other, reserving the label ‘relation’ to indicate the specific way in which the terms are related. For example, ‘the teapot is on the table’ is a relational situation, the ‘on-ness’ being the specific type of relation involved. It is Anderson’s thesis that in order for terms to become related to each other, they must first individually possess qualities or properties of their own, that is, they must have existence that is independent of the relational situations into which they enter. Failing this condition, there could be no relational situations, because there would exist no terms to enter into them, “… unless things had qualities of their own, there would be nothing to have relations to other things” (Anderson, 1930, p.43). Therefore, it is only by existing prior to entering into a relational situation, by having qualities of its own, that a term can enter into a relational situation with another term, which likewise must have qualities of its own.

Furthermore, a term’s participation in a relational situation is not to be confused with its qualities, as if, in and of itself, coming to stand in relation to something else added to the number of a term’s qualities, while ceasing to participate in a relational situation

---

20 The third variation on symmetry in relations, ‘non-symmetry’, receives special attention from Anderson (1928), who takes care to note that ‘non-symmetrical’ is being used in a technical, not a logical sense; the logical sense of ‘non-symmetrical’ including all relations which are not symmetrical (i.e., asymmetrical as well as non-symmetrical relations).
resulted in the loss of a quality. To the contrary, nothing is constituted by the relational situations into which it enters. In Anderson’s (1930) words, “To say that X has the relation R, or has the relation R to A… is not to say what sort of thing X is, and, if the above were all the information we had about X, we should not know X at all” (p.43). Accordingly, terms are to be understood as constituted by their qualities, not by the relational situations they enter into.

Putting these points together we see that, “In the proposition ‘the term a is in the relation R to the term b’, aR in no degree constitutes b, nor does Rb constitute a, nor does R constitute either a or b” (Marvin, cited in Anderson, 1927a, p.27). Returning to the above example, the situation whereby ‘the teapot is on the table’, we see that ‘the teapot is on’ does not constitute the table, ‘on the table’ does not constitute the teapot, and ‘on’ constitutes neither teapot nor table.

Understanding, then, that terms exist independently of any relational situations they become involved in, there remains the question of whether relations, too, can exist outside of relational situations. The answer to this question can only be a decided negative, as seen when Anderson (1962) reminds his readers that, “…what must specially be emphasised… is the fact that “a relation” is not a term…” (p.149). Relations, in other words, are not independently existing conglomerations of qualities, but manners of connection. When terms cease to connect in certain ways, it is not the case that the relation between them lingers until needed at a later time, it simply ceases to be. Relations, therefore, are only to be found in relational situations, and even there not as additional entities besides the terms that are related, rather as the terms’ manners of connection. When a teapot is lifted from a table, the situation ‘the teapot is on the table’ no longer obtains, and while both the teapot and the table continue to enjoy their respective existences, the ‘on-ness’ is at once dissolved. To treat it otherwise, as a term in addition to the referent and relatum, or as a quality of one or the other of them, would be to reify it (see McMullen, 1996).  

21 The failure to recognise that a relation is not an entity creates all manner of confusion as demonstrated in a paper by Daniel von Wachter (1995), who rightly claims that it is difficult to construe a connection between terms and relations if the relations, like the terms, are assumed to be distinct entities, and then concludes from this that “… one can do without external relations” (p.355). A better conclusion to have drawn would have been that relations are not entities at all, and therefore that the question of how they are connected to terms is ill formed.
Needless to say, there exist numerous types of relational situations, each with its own peculiar characteristics, but each with characteristics in common with the others. The independence of the terms that enter into relational situations with each other, the facts that each of these is comprised of qualities and that none are constituted by the relations they enter into, and the non-existence of relations outside of relational situations, are some of these common characteristics; knowledge of them prepares the way for some more specialised considerations that are of importance to Direct Realism, that of causal relational situations (Section 4.3) and cognitive relational situations (Section 4.4.2).

4.3 Causal relational situations

4.3.1 The referents of causation

As an introduction to his remarks concerning causation, Aristotle (1961) writes that “…men do not think they know a thing till they have grasped the ‘why’ of it (which is to grasp its primary cause)”. Agreeing that it is desirable to understand natural change, Aristotle proposes four different ways in which we approach the why, rather than the that, of an effect’s occurrence, in other words, four different conceptions of what it is to be causal. The four suggestions can be elaborated as follows,

The material cause of something is the substance from which it is made, the substance which endures despite any alterations in appearance that it suffers. For example, in the case where a silver bowl is fashioned, the silver itself is the material cause.

The formal cause of something is its archetype, or the definition of its essence. The form of the silver bowl, for example, involves the features, such as roundness, by which bowls are defined.

The efficient cause of something is that which ‘makes of what is made’, for example, the manipulation that changes the material (in this case the silver) into the form (that of a bowl).

The final cause of something is ‘that for the sake of which a thing is done’. Thus, the completed silver bowl is a cause of the production, or efficient causation, of itself.
Of these four Aristotelian conceptions of the character of causes, Anderson (1938) argues that the first two are better understood as characterising the effect than the cause, and that it is the third characterisation, that of efficient causes, that is best suited to a realistic approach to causation (final causes being rejected outright for reasons to be considered below). He writes that, “Thus the effect, in ordinary usage, corresponds to the “formal cause” in the Aristotelian classification – at any rate, to the acquisition by the “material cause” (which, here, is the field or the relevant part of the field) of the form or character in question. And, “the matter having the form” being actually the effect, we are left with the “efficient cause” as the cause proper – which is still in accordance with ordinary usage” (p.133). Accordingly, ‘causation’, as characterised by Anderson, is to be understood as ‘efficient causation’ that is, as a relation in which the ‘cause-term’ brings about the effect.

That causation involves change, and that the cause-term in causation is the active factor in this change, together suggest that the referents of causation are events. But could it be the case that static situations can be causal? In other words, is it feasible that standing conditions might be causally efficacious in the production of effects? To this question, some authors reply in the affirmative.

Fred Dretske (1988), for example, argues that there exist two legitimate ways of conceptualising causes: firstly as triggering events and secondly as structuring conditions. He writes that, “In looking for the cause of a process, we are sometimes looking for the triggering event: what caused the C which caused the M. At other times we are looking for the event or events that shaped or structured the process: what caused C to cause M rather than something else” (p.42). Consider Dretske’s example whereby a wire is soldered in such a configuration that, when a button that turns on an electric current is pushed, a certain effect M (a light’s going on) is caused rather than an alternative effect N (a bell’s ringing). The wire’s being soldered in this configuration shapes or structures what the depression of the button causes, the light going on rather than the bell ringing. Should the wire have been soldered in a different configuration, then the depression of the button could have resulted in the bell ringing rather than the light going on. What Dretske identifies in the pushing of the button and the configuration of the wiring respectively, are two types of causes. Of these he writes, “The first type of cause, the triggering cause, causes the process to occur now. The
second type of cause, the structuring cause, is responsible for its being *this process*, one having M as its product, that occurs now” (p.42).

In other words, according to Dretske (1988), triggering causes are events that initiate causal processes, whereas structuring causes are conditions that determine the directions causal processes take, the effects to which they lead. Seeking the cause of an effect can therefore either involve discovering why the effect occurred when it did, in which case an *event* which triggers the causal process will be sought, or how the cause brought about this, rather than another, effect, in which case a structuring *condition* will be sought. By this analysis, both events and conditions can function as causes.

That similar causal events can produce differing results, such as when on one occasion pushing a button turns on a light whereas on a second occasion pushing the same button rings a bell, certainly demands explanation. But the correct explanation does not necessarily reside in the conclusion that because different effects require different causes, something more than a triggering event is required as a cause. Rather, what needs to be recognised is that something crucial has been left out of this formulation. In the passage that was quoted above, Anderson referred to the ‘material cause’ as ‘the field’, this latter being an integral although oft-neglected feature of causal relations. The field that Anderson refers to is that which the cause acts upon and wherein the effect is produced. In the example involving the production of a silver bowl, the silver was acted upon, perhaps through heating and hammering, in such a way that an effect was produced, a bowl was fashioned. Should a different field have been acted upon by the same efficient cause, for instance, had the field been a block of wood, then a different effect would have resulted. The fact that the same efficient cause can produce different effects if introduced to different fields is, Anderson argues, to be expected, and does not provide evidence that there is anything inadequate or unlawful about the connection between causal events and their effects. Rather, “… a cause is always a cause within a field… A may be necessary and sufficient for the occurrence of B within the field X, and yet not be necessary or sufficient for its occurrence within the field Y. And the fact that A cannot, as we say, make a Y become B, is nothing against its having

22 Note that ‘field’ is being used differently here from above (Section 4.2), where it indicated the domain and converse domain of a relation. The present usage is meant to indicate not the class of all possible causes and effects, but the class of all causally relevant conditions that a cause interacts with to produce an effect.
that effect on an X and suggests no variability in the causation of B in the field X” (Anderson, 1938, p.130).

To some extent, Anderson and Dretske are in agreement. As Dretske (1988) claims, “an event of type C causes or brings about an event of type E only in a certain restricted or special set of conditions. Call these background conditions. If the right background conditions do not obtain, C will not cause E. You can push the button all you like; the bell won’t ring unless the wires are connected. Once the wires are connected, pushing the button causes the bell to ring” (p.39). Dretske’s background or structuring conditions are Anderson’s causal field. When they are of one sort (X), a cause (A) brings about an effect (B), and when they are of another sort (Y), the same cause (A) does not bring about (B). There remains, nonetheless, an important point about which Anderson and Dretske disagree, and this concerns whether or not causal field conditions are the referents of causation. In accordance with Anderson, I would suggest that standing conditions, Dretske’s ‘structuring causes’, are best understood as the field into which causal events are introduced, and within which effects are produced, in other words, that the same causes produce the same effects within the same fields. This leaves, as the candidates for causal referents, those events which impinge upon the relevant field and in so doing alter it.

4.3.2 The relata of causation

The relata, or effects, of causal relations have received much less attention than have causal referents, perhaps because the question of what it is to be an effect is much less controversial than is the question of what it is to be a cause. Nevertheless, as we have seen, Anderson (1938) characterises effects as the field’s (Aristotle’s ‘material cause’) coming to acquire a certain character or quality (Aristotle’s ‘formal cause’). This may suggest that causal relata, too, are events, but Anderson refines his characterisation thus, “… the position is that a member of the genus (or part of the field) acquires a character which it previously had not; and it is this acquisition, or, more exactly, the thing’s now having the character, that we speak of as an effect. For example, when something makes me blush, it is “my blushing now” that is said to be the effect of its operation, and not my transition from non-blushing to blushing, though it has to be understood that I was not blushing before” (pp.132-133). Thus, the relata of causation are comprised of newly acquired characteristics of causal fields. Returning to the previous example, the silver’s
being bowl-shaped is the effect of the efficient cause (heating and hammering) impinging upon the field (the silver itself).

4.3.3 Causal necessity, sufficiency, laws and priority

Acknowledgement that causation always take place in a field is the key, Anderson (1938) argues, to resolving ambiguities that arise with regard to questions concerning whether causes are necessary, sufficient, both or neither, for their effects, and, accordingly, whether causation is lawful or suffers from ‘variability’, that is, exceptions. That causes are both necessary and sufficient for their effects, within a field, and that variability is at odds with causation, are, Anderson argues, both facts without which, “… we could not say that there was any causal connection at all” (p.126). This is because, in endeavouring to discover causes, “… in distinguishing the relevant from the irrelevant, that is to say, the necessary from the unnecessary, we are concerned with general conditions (necessity being equivalent to universality), and, if we do not find a general condition of a given occurrence, we are not answering the question that has been raised” (p.127). Thus, while Davidson, for example, argues that causation occurs between particular events (that is, events themselves independently of any universal properties that may be attributed to them), Anderson insists that, “When we ask, for example, what causes this fire, it is not its being this but its being fire that we are seeking to account for… it is fire, a certain sort of thing, that is the effect in question, and, if any distinction is to be made among conditions of its production, it will be a distinction between different kinds of conditions. It is natural, then, that, to the question what causes a certain sort of thing, the answer should be “a certain sort of thing”; it appears that what we are all the time seeking to establish is a general connection, that is to say, a universal proposition, to assert which is to assert that something happens invariably” (p.127).

Anderson’s reference here to what can be asserted of causal relational situations should not be mistaken as suggesting that he, like Davidson, believes that, “In fact, it is not events that are necessary and sufficient as causes, but events as described in one way or another” (Davidson, 1969, p.172). Rather, Anderson is arguing that certain sorts of causes are necessary and sufficient for the occurrence of certain sorts of effects within certain sorts of fields, that is, that causation obtains between types of events and conditions in the world. Because they obtain in the world, causal regularities can, at
times, be *discovered*, and when this happens, they can then become the objects of causal explanations, that is, be asserted to exist. But, of course, such discovery is possible only when the regularities occur in the world, these occurrences being independent of human observation and description.

So, what is meant by saying that a cause is necessary and sufficient for its effect within a field? This question has been considered by Anderson’s student John Mackie (1980), with particular attention paid to the role that conditional analyses can play in elucidating what, precisely, one means by claiming that a cause is necessary and sufficient for its effect, within a set of circumstances (a causal field). Of a sequence whereby *X* causes *Y*, he writes that causal *necessity* can be understood to mean that, “… in the circumstances *Y* would not have occurred if *X* had not” (p.31), and that causal *sufficiency* can be understood to mean that, “… given the circumstances, if *Y* had not been going to occur, *X* would not have occurred” (p.39).23

Necessity and sufficiency have been treated, thus far, insofar as they are relations that proceed from cause to effect, but, as Anderson (1938) indicates, “… when a cause is taken as a necessary and sufficient precedent condition of the occurrence of a phenomenon (its “effect”) in a certain field, then it follows that the effect is a necessary and sufficient *subsequent* condition of the occurrence (or operation) of the cause in the field” (p.131). This is because “… if *A* is necessary for *B*, *B* is sufficient for *A*, and *vice versa*” (p.131). In other words, ‘*A* is necessary and sufficient for *B*’ is a symmetrical relational situation, it implies that ‘*B* is necessary and sufficient for *A*’. Despite this, causation is an asymmetrical relation, if ‘*A* caused *B*’ then ‘*B* did not cause *A*’. Consequently, there must be some further feature present in causation that accounts for the asymmetry, something that Mackie (1980) denotes ‘causal priority’. As may be gathered from the above, Anderson deems this priority to be temporal, the cause is a ‘necessary and sufficient *precedent* condition’ of the effect, while the effect is a ‘necessary and sufficient *subsequent* condition’ of the cause.24

---

23 Mackie’s investigation involves a much more in-depth and interesting analysis of sufficiency and, in particular, necessity, than is indicated here, but consideration of these details would surpass present purposes.

24 Mackie, on the other hand, denies that causal priority is equatable with temporal priority, arguing instead that, “… temporal priority is not conceptually necessary for causal priority” (p.52). Mackie’s argument turns not on any evidence that backward causation, that is, causation whereby the effect precedes the cause, actually occurs, rather, on the invaluable role he sees ‘fixity’ playing in our concept of causal priority. He argues that, “… if at any time *A* is fixed while *B* is still unfixed, *B* cannot be causally
Causal priority, particularly when understood as an instance of temporal priority, provides a hint as to why the notion of ‘final causes’ is at odds with determinism. A final cause, such as a finished silver bowl, is postulated to be efficacious in the causation of events that not only occur prior to it, but that are ordinarily understood to be efficient causes of it, in this example, events such as heating and hammering. But, if the temporal priority reading of causal priority is correct, then causes are necessary and sufficient precedent conditions of their effects, and of these, B, being temporally posterior to A, cannot be a part.

Additionally, if the proponent of final causation does indeed admit that what the final cause causes is, at the same time, an efficient cause of the final cause, that is, that B is a final cause of A and that A is an efficient cause of B, then he or she is proposing that each is a cause of the other, and, accordingly, that each has causal priority over the other. This, of course, involves a degree of circularity which makes nonsense of the asymmetry involved in causation, that is, the fact that if A causes B, B does not cause A. It is for these reasons that the notion of ‘final causation’ cannot be upheld.

4.3.4 Causation and mental events

Finally, in characterising causation, Anderson stresses that it applies uniformly to mental and non-mental events, that there is no question of ‘special causation’, such as goal-directed or willed action, that differentiates human activity from that of the doings of non-human objects. He writes that, “If, then, there is a “categorial law” of causality, there will be no question of its applying differently to mental and to non-mental events, no question of “kinds of causality” (e.g., moral determination and physical determination), though there will certainly be a question of kinds of causes – one event may have or be a cause which is of a moral quality, and another may have or be a cause prior to A, because at this time things might still be so decided that B does not occur” (p.178). It is feasible, Mackie believes, that a temporally posterior item B could become fixed before a temporally prior item A, thus providing evidence for backward causation, if the following situation were to obtain. Item A occurs at time t2, item B occurs at time t3, and item C, which is sufficient for the occurrence of item B, occurs at time t1 (whereby t1 precedes t2 which precedes t3). If this situation were to obtain, if C were truly sufficient for the occurrence of B and if A were not fixed until it actually occurred (i.e., if A were unfixed until t2), then the ‘fixity’ condition would be met, and this would remove any inconsistency that might arise from the claim that ‘B caused A’ when it is supposed that B might yet be prevented from occurring. But, as Mackie warns, the ‘fixity’ criterion of causal priority rests on the supposition that there are, in the world, certain occurrences that only become ‘fixed’ when they occur, in other words, occurrences that do not have sufficient precedent conditions; and this possibility is incompatible with determinism. Acceptance of the ‘fixity’ criterion of causal priority at the expense of the temporal criterion requires, therefore, the adoption of indeterministic principles. For this reason it must be rejected by any, such as Direct Realists, for whom determinism is a given.
which is of a physical quality, leaving aside the question whether what is of a moral quality is not also of a physical quality” (Anderson, 1939, p.263).

Anderson’s reason for denying mental events a special status, either as uncaused or as causing other events in some way other than efficiently, derives from his recognition of the fact that people are part of nature, and that they are, therefore, subject to its ways of working. Furthermore, he demonstrates that if, as indeterminists typically wish to maintain, undetermined mental events interact with determined physical events, then there could be no such thing as lawful physical interactions, because, “… any physical “uniformity” is to the effect that a certain set of physical antecedents gives place to a certain set of physical consequents; but, with the intervention of a free agent, the very same set of physical antecedents will have a different set of physical consequents. It is surely clear that if, with no physical difference in the antecedents, there is a different physical sequence, then there can be no physical uniformity. This is simply an illustration of the impossibility of combining the free and the determined in any situation” (Anderson, 1936, p.124). The counter-position to this interactionist one is that, “… we have the same ground for treating man as necessitated as we have for so treating anything else” (Anderson, 1939, p.260). This, then, is the premise from which a Direct Realist account of mental causation begins.

4.4 Maze’s Direct Realist theory of mental causation

The argument for a Direct Realist approach to cognition and mental causation has been developed at some length by John Maze, and appears in his book entitled The Meaning of Behaviour (1983). Unlike Davidson, Maze aims to provide a thoroughly deterministic account of mental causation, that is, one that contains no attempt to reconcile human freedom with natural necessity. Accordingly, he differentiates between causal and teleological explanations of behaviour, and rejects the latter because of their contained assumption of organismic agency, which, he argues, “… is a concept that is simply unintelligible, one that cannot offer any useful information” (p.12). Maze’s alternative to the notion of self-directed behaviour is that there exists a plurality of primary drives within higher organisms, including humans, whose processes motivate and sustain behaviour until such time as they become deactivated. This mechanistic account of behaviour is aimed at demonstrating how what may appear to be goal-
directed or purposive behaviour is, in fact, the effect of efficient causation, that the occurrence of behaviour, in other words, is simply a species of natural causal process.

In addition to its mechanistic stance, Direct Realism differs from Davidson’s and, indeed, most other extant theories of cognition, because it rejects the view that mental phenomena are internal bodily states, proposing instead that they are external *relational situations*. The claim that cognition is relational, while novel to many, is relatively easy to accept when it is recognised that cognition is the occurrence of an organism coming to know something about an object of thought, in other words, of an organism coming to *relate* to an object of thought in this special way. What can be less easy to come to terms with is the fact that the objects of thought, being external to the cognising organism, necessitate that the cognitive relational situation, too, is external to the cognising organism. Indeed, resistance to the external character of cognition through the postulation of internal, surrogate objects of thought has been actively maintained from the earliest known psychological theories, those of the pre-Socratic Greeks (Hospers, 1956), throughout the Middle Ages (Pasnau, 1997), and continues without abate in present-day accounts of ‘mental representations’. Nevertheless, as will become evident, repugnance to the external character of cognitive relational situations must be overcome if a viable, realistic theory of cognition, and with it mental causation, is to be developed.

In what follows, I shall consider firstly the role of drives in Maze’s theory (Section 4.4.1), then his characterisation of cognitive relational situations (Section 4.4.2), and finally his account of mental causation (Section 4.4.3).

**4.4.1 Drives**

As already indicated, at the heart of the Direct Realist position is a strong commitment to determinism, the thesis that, “… every event has a cause, and is a cause of further events” (Maze, 1983, p.8). The acts of human beings, whether cognitive or behavioural, form no exception to this rule, and hence the folk psychological conceptions of free will, self determined acts and goal-directed or teleological behaviour are fatally flawed because they assume that human beings can choose if, and how, to think and behave, with at least some degree of independence from necessary and sufficient causal circumstances and events. Consider once more William James’ suggestion that he is free to choose which route to take home from his evening lecture, and, his assertion that
if only the clock could be turned back, and all other circumstances remain the same, he
would then be free to choose otherwise than he had originally. The substance of James’
assertion is that, unlike the inanimate objects in the world whose every change is fully
determined by prior-occurring causal events, human beings have the capacity for free
choice, which is to say, to think and behave indifferently to the causes that impinge
upon them. It is Maze’s aim to dispel the illusion of free human action that is advocated
by James and others by elucidating the mechanisms by which thought and behaviour are
determined. This endeavour focuses particularly upon providing a deterministic account
of the apparent goal-directedness of the behaviour of higher organisms. Part of the
mechanism by which thought and behaviour are determined is postulated by Maze to
consist in an as yet unknown number25 of feedback controlled primary drives that
consist of biological or physiological engines and innately provided specific actions. An
outline of Maze’s drive theory follows below.

As emphasised by Maze (1983), the biological engines that drive behaviour are
physiological structures which, once identified, will be open to description in terms of
their intrinsic physical characteristics. There is nothing metaphysically suspicious or
mysterious about these structures, indeed, the concept of a biological engine is “… as
literally the concept of an engine as the concept of a car engine is” (p.136). Maze argues
that, like a car engine, a biological engine is not a source of energy, nor does it engage
in causal relations of a type different from that of anything else. Rather, inputs from
various sources activate the biological engines which respond in ways that typically
involve bodily movements. These movements consist, at first, of innately provided
actions and later, of more sophisticated actions that derive from the simple innate ones.
The kinds of movements that are made typically, although not always, result in the
cessation of the input conditions, and thus the activity of the biological engine, and thus
of further movement too. For example, certain chemical and somatosensory events can
lead to the activation of the biological engine that is a component of what is known as
the hunger drive, causing an infant to make sucking and swallowing movements. If
these movements result in the intake of sufficient quantities of food, then the initial
conditions that activated the engine will no longer persist, the activity of the engine will
cease, and the sucking and swallowing movements will stop too. Under these

25 Drive structures should not, Maze’s insists, be postulated to be infinite in number (in the way that
beliefs and behaviours can be), rather, they are to be identified by their somatic sources, that is, by their
physiological make-ups.
circumstances, that is, when reversal of the initiating conditions occurs and engine activity and movement cease, the movements that altered the initiating conditions are said to be *consummatory*. In the present example, the sucking and swallowing movements that result in food intake are consummatory actions (see Maze, 1983, p.153).

Thus far, we have seen that the primary drives are feedback controlled mechanisms consisting of biological engines and the simple innate actions to which the engines give rise. In Maze’s (1983) words, “The concept of drive, then, must include not only that of an internal mechanism which when activated impels the organism to action, but also that of *the innately provided specific actions which it impels, and whose performance is a necessary condition of the termination of the drive state*” (p.142). From these humble beginnings, the complex doings of human beings arise. Indeed, it is Maze’s thesis that, “… *everything one does throughout one’s life, however obviously acquired, sophisticated or culture-bound it is, is some modified form or instrumental elaboration of one of the innate consummatory actions*” (p.152). The step from biological engines impelling simple innate actions to those same engines impelling sophisticated behavioural acts may seen to be a large one, but it is a plausible one nonetheless. Maze argues that what has appeared to many to be either goal-directed or self-directed behaviour is in actual fact the product of the fine tuning of the movements that the engines compel; fine tuning in accordance with the structure of the environment that is interacted with. He considers as an example of this the elaboration and refinement of the eating movements engaged in by a human as it passes from infancy to adulthood. At the beginning of this developmental journey, the eating programme is quite simple, “In the beginning it consists of this: if the infant has been deprived of food for a sufficient period, then when food is presented to it, *it eats the food*… The output side of the program, eating, at first consists of just the complicated interaction of lips, tongue, jaws and throat that constitutes sucking and swallowing” (p.68). Furthermore, these simple actions are innately provided, “No one who has watched an infant’s first effort at sucking and swallowing could seriously hold that those component movements need to be assembled by learning, it seems to me” (pp.68-69).

However, innately provided simple actions such as sucking and swallowing will only suffice as consummatory actions for so long, and therefore if this were all that the
activation of the biological engine gave rise to, the organism would soon perish. Fortunately, learning allows for the elaboration and refinement of the eating programme, in terms both of the inputs (what is to count as food) and the outputs (how the organism’s body comes to move). Maze (1983) writes that,

The effective input or stimulus aspect of the program becomes elaborated as the young organism discovers various facts about food – most importantly, where it is and what needs to be done to get it. As this happens, those other types of object which are discovered to be regularly and understandably related to food come to act, in a sense, as *surrogates* for food when the baby is hungry and actual food is not present. I say ‘surrogates’ in order to stress the fact that they *cause* the eating program to begin running off, whereas in the first place it was only food itself that would do so. In another way it would be better to say that the associated objects stand as *signs* of where food is, or perhaps that the stimulus ‘food’ has become a complex that includes the associated objects, since the young organism knows them propositionally as related to food. In any case, if an associated object causes the eating program to begin running off in the absence of actual food, the program would obviously be abortive if its motor output continued to consist only of sucking or chewing and swallowing – that is, if the response to the associated object were the same as the one to the original stimulus, food (p.69).

What should *not* be concluded from the fact that an alteration in bodily movements is necessary if eating is to be successful is that such an alteration will inevitably occur. Maze (1983) stresses time and again that *requirement* is not a causal force, that the mere fact of A’s being needed for B to occur is irrelevant to whether or not A will happen. He continues,

Instrumental actions need to become incorporated in the eating program if it is to get food into the stomach. As I have been saying, the need for them does not mean that they *will* be incorporated; to retain the determinist view we must suppose the genetic provision of the coordinating nerve centres that I hypothesised above, and certainly the very early age at which some of these instrumental acts begin to be incorporated makes that a plausible supposition. One very early elaboration of the eating program is for the baby to turn its head so as
to bring its mouth to the nipple. Later comes the grasping of objects and transporting them to the mouth, and then progressively the various forms of bodily locomotion – shuffling, crawling or walking to the food, before grasping it and putting it into the mouth (p.69).

Already it can be seen that, by the time an infant can crawl or walk, the innate actions which the biological engine involved in eating impels have been greatly refined and elaborated. At this point they are still recognisable as elaborations, however, as their complexity increases, their recognisability tends to diminish, and by the time that adulthood is reached, their disguise is effective to such an extent that the illusion of goal-directed or self-directed behaviour can arise. Maze continues his explication thus,

Allowing a span of years to elapse, consider the following everyday example of the elaboration of both input and output in my eating program. I am sitting in my room at the university and gradually become hungry. I know that the nearest source of accessible foodstuffs is the cafeteria in the Students Union building. I know the geographical relatedness of my room and the cafeteria – the passage outside my door, the stairs, the exit from this building, the path beneath the plane tree, the door of the Union. Because of that information, because of my state of drive, because there are no other programs running off in me at the moment that prevail over the eating program, the sight of the door of my room causes me to get up and walk through it, the familiar passage outside is responded to as path-to-stairs-and-food rather than, say, path-to-lecture-room, and so on, until I come to the food. Each succeeding part of that path is seen by me in its relation to food (though this is not to deny that I also had the conception of the whole path before I started), and so keeps my extended eating program running (pp.69-70).

The above illustration, Maze acknowledges, is consistent with the S-R version of behaviourist psychology in that it concedes that behaviour is always a response to the environment, but at the same time it differs in at least one important and crucial respect from behaviourism, this difference being that while behaviourism denies cognition a causal role in behaviour production, Maze insists upon it. He continues,
… one may think of each of the various food locations with which I am familiar as having a great number of invisible tentacles spreading through the neighbourhood, each tentacle being a path that I know of from some place to that food location, so that wherever I am when I get hungry, provided I know even in general terms the spatial relation between the place where I am and some food supply, that present stimulus situation causes me to start moving away from it in the direction of food (p.70).

In the above passage, Maze carefully avoids giving the impression that the incorporation of cognition into the drive programme enables the organism to wilfully alter its bodily movements. The point is more explicitly argued for in what follows,

Further, although it is a strained usage, what I can most informatively be said to be doing, even as I take the first steps towards the food, is *engaging in eating*. That is to make the point that it is still the same eating program which in its germinal form I acquired genetically that is running off in me, and it shows also that the language of purposivism, which would say that I am walking *in order* to get the food, is not inescapable. It is possible to express in a deterministic way the relationship between the instrumental activity and the consummatory activity that motivates it by saying that the former has become incorporated as a *part* of the latter in an extended and elaborated version of the same basic program… Thus, the eating programme for most adults in industrial communities must include exchanging money for the food, so that the occupations one takes up for the receipt of money also count as part of the eating program, although of course they will serve as part of other programs as well (p.70).

In summary, primary drives are simply innately provided physiological mechanisms that are activated by sensory inputs and deactivated by the cessation of such inputs, in a thoroughly deterministic manner. The cessation of input stimuli is typically brought about through the intervention of appropriate bodily movements or *consummatory actions*, and it is the refinement of the functioning of the primary drives, or the fine-tuning of these consummatory actions, that accounts for the shaping of what has appeared to many to be goal-directed behaviour. Such shaping, or *elaboration* (as Maze refers to it) is mediated by cognition, that is, “… through the formation of beliefs about
the locations of the objects on which the consummatory actions are performed, and about the likely effects of actions on objects in the present environment and the relation of those effects to the getting of the goal objects” (Maze, 1983, p.70). Given the central role of cognition in a Direct Realist account of behaviour production, it is necessary to consider this relational situation in more detail.

4.4.2 Cognitive relational situations

As already mentioned, Direct Realism postulates cognition to be a type of relational situation, indeed, a non-symmetrical binary relational situation involving a knower (the a term or referent) and what is known (the b term or relatum). Importantly, these terms are taken to exist independently of each other, and independently of any relational situations, including cognitive relational situations, into which they enter. The present section outlines Maze’s (1983) account of the referents (4.2.2.1), relata (4.2.2.2) and relational character (4.2.2.3) of cognition.

4.4.2.1 The referents of cognition

The domain of cognition is comprised of any and all entities that can enter into cognitive relational situations. Maze (1983) characterises individual referents of cognition, and distinguishes them from pseudo-cognitive machines, by the following criterion, “An entity that really can know is one whose sense organs can scan the world and discriminate and register things of a kind it has never seen before…” (p.81). The kinds of entities that are at present capable of cognising include only organic ones, but Maze concedes that, despite their absence at present, “… genuine knowing machines may be built in the future” (p.81).

The distinguishing feature that Maze (1983) takes to define the domain of cognition is the capacity for genuine learning, that is, the ability that some entities have for changing from a condition of ignorance to one of awareness, through sensory contact with the world about them. As an example of an obvious candidate for a referent of cognition, Maze cites a human infant who, “…coming into the world almost completely ignorant, and certainly with no articulated visual experience, must be able natively to see (cognise) kinds of thing that it has never seen before and re-cognise them when they appear again” (p.81). This capacity to change from ignorance to awareness is what discriminates cognising entities from those that do not cognise, that is, those that do not
and cannot be aware of anything at all, as is the case with present day computers that can at best mimic cognitive processes to the finite extent that their programmes allow.

As may be gathered from the above, establishing the criterion for the class of possible referents for cognitive relational situations is simply a matter of definition. Of more difficulty is identifying the individuals who meet the criterion, and, more particularly, identifying what structures within those individuals are responsible for their inclusion in this class. These latter tasks are empirical matters that inevitably lead to thorny issues. For example, at what point, if any, on the phylogenic scale, do animalistic organisms cease to meet the criterion? Davidson’s ready answer to this question, that only language-using animals can think, is inappropriate because cognition as here defined is a non-linguistic activity, one for which language is unnecessary. Can plants cognise? They certainly react differentially to their environments, but is this what is meant by discriminating and registering things about the world? An answer to these questions will be attempted in Section 4.4.3, but for now it suffices to note that identifying the domain of cognition is a difficult task, and one that could well involve a great deal of open-mindedness.

What about the second empirical matter, identifying the material structures that are responsible for an entity’s capacity to scan, discriminate and register facts about the world? This task is of an even more difficult nature than the first, and presupposes the resolution of the first (clearly one must first identify which entities can cognise before beginning to speculate about how they do it). Nevertheless, it is not a hopeless enterprise, particularly so because some cognising entities have been identified, and of these one can make inquiry as to which bodily structures are relevant for the occurrence of cognition (recognising meanwhile that other cognising entities might make use of different types of structures). Maze’s (1983) attention is largely focused upon human psychology, but clearly his hypotheses may be relevant to non-human species that share with humans similar bodily structures.

Considering, then, cognising entities such as human beings and their cousins, Maze (1983) argues that the structures enabling such organisms to cognise are intrinsic brain states. To this end he comments of the cognitive relational situation that, “…when it is a relationship peculiar to certain kinds of entity (higher organisms), one would expect to
find also that those organisms had particular *intrinsic* properties in virtue of which they stood in the relations in question. Since cognitive relations (the beliefs an organism has) are so particularised and variable, then it seems reasonable to assume that any instance of such a relationship, i.e. a belief in a particular fact, would be subserved by an intrinsic state of the organism specific to that belief, and the obvious kind of intrinsic state to look for is a brain state” (p.86).

In attending to the relevance of brain states to the occurrence of the cognitive relational situations engaged in by higher organisms, Maze (1983) is advocating a mind-brain identity thesis, but one that differs greatly from that suggested by Davidson, and, indeed, one that is very different from the majority of mind-brain identity theses that are currently endorsed. Contra Davidson, brain states are not to be conceived of as simultaneously ‘inner’ and ‘outer’, depending on whether their descriptions are in physical terms or mentalistic terms, rather they are purely ‘inner’ or intrinsic in the sense that they are contained within the cognising organism’s body.

Furthermore, the brain states that underlie an organism’s cognitive relational situations are not to be thought of as intentional, that is, as containing meaning or as representing anything about the world, rather, “Any such specific brain state would simply be whatever it physiologically is; to call upon proposals that have been made, though none of them now seems technically feasible, it may be one of Hebb’s (1949) ‘cell assemblies’, or a ‘reverberatory circuit’, or some specific ‘tuning’ of certain neurones, but whatever it may physiologically be, it would in principle be describable in purely physiological terms without any reference to the cognitive relationship it underlies” (Maze, 1983, p.86).

The importance of recognising the independence of brain states from the cognitive relational situations that they enter into should not be underestimated, nor, as was seen above, is such independence a fact that is peculiar to these types of referents and this type of relation. As argued above (Section 4.1), no referent is comprised, even partly, by the relational situations it enters into. With more particular reference to cognitive relational situations and cognising organisms, it needs to be understood clearly that in entering into cognitive relational situations, brain states do not take on as properties these situations, that these relational situations remain extrinsic to them, and,
accordingly, that the cognitive relational situations cannot be discovered within them any more than a spatial relationship to a table can be discovered within a teapot (recall the earlier example of the relational situation, ‘the teapot is on the table’ in Section 4.2).

In this way, Direct Realist theories of cognition can be seen to differ from most mind-brain identity theories in that they reject the notion that the relevant brain states are something more than bodily states, something, more specifically, that provides reference to or represents the relata which supply the second term of the relation and that can be discovered by examination of the state itself. These negative points are emphasised by Maze (1983) as follows, “… it is logically impossible that a physical entity such as a ‘cell assembly’ (or whatever physiological state it may be) could have intrinsic to it, and thus discoverable by inspecting it, its relation to something external to it, the fact believed in” (p.87).

Consider, for example, the circumstance whereby Helga cognises the fact that ‘this apple is sweet’. The referent in this relational situation is Helga, or, more specifically, a brain state of Helga’s, and the relatum is the situation of the apple’s being sweet. The point that Maze is emphasising is that at no stage does Helga’s brain state contain any reference to, or representation of, the apple’s being sweet. Consequently, it is in-principle impossible to discover by examination of Helga’s brain state what it is in relationship with, or, as it is usually termed, what the content of the cognitive relational situation is. This impossibility, it must be emphasised, does not arise through technological limitations that might be overcome in the future, rather, it is a consequence of the fact that no referent contains its relations, nor what it is related to, within it.

Thus far, we have seen that while the domain of cognition is difficult if not impossible to enumerate exhaustively, it is at least possible to point to examples of referents of cognition, such as humans, and to ask of these which of their physiological structures, such as their brain states, become involved in cognitive relations. There are thus more and less refined answers available to the question of what forms the domain of cognition. A less refined answer suggests that at least part of the domain is comprised of human beings. A more refined answer suggests that at least part of the domain is comprised of the brain states of human beings. These two answers are compatible with
each other, and also with a third, an answer that is more refined than ‘humans’ and less refined than ‘brain states’, namely, ‘drive structures’. It is this latter that Maze (1983) proposes to be the most illuminating answer to the problem of specifying the referents of cognition, at least as concerns humans, particularly as it accounts simultaneously for the connection between cognition and motivation that is necessary for the occurrence of behaviour. He writes that, “If mental processes are relations into which brain processes enter, as central state materialism proposes, then on the instinctual drive theory it is specifically the drive structures which, through their connections with the perceptual system, enter into those cognitive relations” (p.162).

In summary, the domain of cognition is comprised in part of human cognisers, and the structures which allow for humans to cognise are physiological bodily structures. The identification of the particular bodily structures involved in human cognition is, at present, incomplete, although in gross terms it may be granted that the brain and sensory organs play a central role.

4.4.2.2 The relata of cognition
The converse domain of cognition is comprised of environmental situations which exist independently of being known. These relata are typically referred to in Direct Realist literature as ‘objects of knowledge’ or ‘the known’. More specifically, cognitive relata are propositions, which is to say situations, facts, or states of affairs existing in a subject-copula-predicate arrangement. In Maze’s (1983) words, “Any feasible account of knowing takes it that the objects of knowledge are propositional, which is to say that the most logically primitive bit of information that we can grasp is the attribution of some predicate to some subject” (p.82).

The assertion that cognitive relata are propositional in character needs to be accompanied by some important details if misunderstanding is to be avoided. Of critical importance is the distinction between the present usage of ‘proposition’ and its more common usage, that which applies to a certain type of sentence. By characterising cognitive relata as ‘propositional’, Maze does not mean to suggest that they are linguistic, rather that they are complex units involving logical (not linguistic) subjects and predicates in the form of ‘$a$ is $B$’. For example, in the situation considered above whereby Helga cognises the fact that ‘this apple is sweet’, the logical subject of the
relatum is ‘this apple’ and the logical predicate is ‘sweet’. What Helga cognises is a complex situation whereby the apple is of a certain character, it is sweet.

Maze (1983) argues that nothing simpler than a proposition can be cognised because nothing simpler than a proposition can exist, and, accordingly, that theories that postulate the building up of knowledge from ‘sub-propositional atoms’ are mistaken. It is this that he is expressing by calling propositions ‘the most logically primitive bit of information that we can grasp’. Sub-propositional atoms, Maze argues, “… could not be assembled piecemeal, as a collection of bits, to form a proposition; the relation between subject and predicate is one irreducible fact, and can only be grasped as such” (p.83). Helga, he would argue, cannot know the subject ‘apple’ independently of at least one of its predicates; nor could she know ‘sweetness’ in isolation. Coming to know that ‘this apple is sweet’ is not a matter of separately cognising ‘apple’ and ‘sweetness’ and then somehow combining these two units, it is a matter of grasping the situation, the sweetness of the apple, in its complexity all at once. Accordingly, by calling the relata of cognition ‘propositional’, Maze means to convey not that they are linguistic in character, but that they are complex units of a subject-copula-predicate form.

A further potential source of misunderstanding that requires mention lies in Maze’s (1983) use of the word ‘attribution’ when he speaks of ‘the attribution of some predicate to some subject’. It will be remembered that Davidson uses the word ‘attribution’ to convey an anti-realist sense of the character of properties, that for him to attribute a property to an object or event is to apply a linguistic description to it that is without adequate empirical basis in the object or event itself. This, in turn, requires the attention of an interpreter, and the mind-independence of the property attributed is thereby denied. Maze, on the other hand, like Anderson, insists upon the independence of the relata of cognition from the individuals who know them and the cognitive relation itself. He argues that, “… the objects of mental acts are not constituted even in part by standing in that relation; they are states of affairs that exist independently of their being known…” (p.99). What Maze is arguing for here is the mind-independent reality of propositional situations in the sense of their existence being independent of, and indifferent to, their roles as terms in cognitive relational situations. A corollary of this is that, in order to become known, such situations must already exist, and can therefore in no sense be created by being known.
In summary, then, the converse domain of cognition consists of all propositional situations, facts, or states of affairs that can in principle come to be known, that is, stand as relata in cognitive relations, regardless of whether or not they ever do. Such propositions exist independently of the relations into which they enter with cognisers, and, as such, are neither created nor altered by being cognised.

4.4.2.3 The relational character of cognition

Direct Realists, it has been seen, understand the referents of cognition to be composed of physiological properties, and, it may now be added, to be devoid of mental properties. The dismissal of the notion that thinkers are composed of both mental and physiological properties is at odds with much of psychological theory, whether past or present, but is compatible with the view that cognising is something that organisms do, not something that organisms are. To better understand the distinction, think back to the example where ‘the teapot is on the table’. Resting on the table is something that the teapot does, not something that the teapot is. What the teapot is is a white, rounded, ceramic vessel. None of the things that the teapot does, containing water, leaving a heat impression, breaking and so forth, contributes to what the teapot is.

Similarly, a cognising organism, such as a human being, is composed of various properties, some of which include mass, extension, chemical composition and so forth. In addition to these properties that dictate what humans are, there are things that humans do, run, sleep, eat and joke, to name but a few. One of the things that humans can do is think. Thinking, therefore, is an activity that an organism engages in, it is not a state of its body. But to characterise thinking only in terms of organismic activity is to ignore what is most distinctive about it, the fact that it is always involves a relation to an external situation. It is their relational characters that Maze (1983) emphasises when describing cognitive processes, “By cognitive processes I mean believing, perceiving, knowing that, being conscious of, remembering, and so on; these can coherently be thought of only as relations into which the organism enters, or more specifically as relations between certain specific kinds of bodily process (primarily, brain-plus-sense-organ processes) and things external to those processes” (p.83). Being a relation between some organisms and external situations is, therefore, a distinctive feature of this particular activity, cognition.
But teapots also relate to things that are external to them, and there would be few, if any, who would suppose their teapots to engage in thought. So what we have so far is that cognition is an activity that is engaged in by some but not all worldly objects, and that it is an activity that always involves a relation to an external situation. What more is needed to capture the distinctive character of cognition?

An answer to this question has been suggested by a student of Maze’s and proponent of Direct Realism, Joel Michell. Michell’s characterisation of cognition is foreshadowed in his critique of Maze’s theory that was published in 1988, but it is most clearly stated in his lecture notes, and reads as follows, “… cognition is organismic sensitivity to the propositional structure and content of environmental situations” (Michell, 2001). This definition contains within it the key to distinguishing cognition from other types of sensitivity to the environment, and also an answer to the question posed above concerning the criterion for who, or what, can be a cognitive referent. These issues will be dealt with in turn.

As a species of organismic sensitivity to the environment, cognition is distinguishable from non-cognitive sensitivities by the peculiar characters of its relata. Above, cognitive relata were characterised as environmental situations that are propositional in structure (Section 4.4.2.2). The uniqueness of cognitive relations between organisms and environmental situations appears upon closer examination of what it is to be sensitive to such situations. In an unpublished paper concerning the continuing relevance of Anderson’s philosophy to psychology, Michell defends Anderson’s thesis that, because there is no gap between what is known and reality (or, in other words, because cognition is unmediated by mental representations), the features of propositionality present in all environmental situations are likewise the features that organisms are sensitive to when cognising. These features, according to Anderson, include space, time and the categories of being; the latter being comprised of identity, difference, existence, relation, universality, particularity, number, order, quantity, intensity, substance, causality and individuality (Baker, 1986, pp.98-105).

Considering for illustrative purposes space, time and a limited selection of the categories (universality, number and existence), one can demonstrate that in order for an
organism to have a cognitive connection to a certain environmental situation, it must be sensitive to each of these situational features. Consider again an example used earlier, Helga’s cognition of the fact that ‘this apple is sweet’. Here we see that Helga must be sensitive to space, time and the categories as follows:

Firstly, the fact that Helga must be sensitive to the spatio-temporal location of the situation is apparent if she is to connect with the subject, ‘this apple’, and be able to discriminate it from other subjects, such as other apples. Secondly, Helga must be sensitive to the predicate of the situation, the apple’s sweetness, and be able to discriminate it from other of its properties, such as its colour and size. This involves sensitivity to the category of universality because each discernible property is universal (i.e., either present or possibly present in other situations). Thirdly, Helga must be sensitive to the category of number as evidenced in the present example by her knowing that this apple (i.e., one apple) is sweet. Fourthly, Helga must be sensitive to existence (or occurrence), that is, to the fact that not only are there present both an apple and sweetness, but that this apple is sweet.26

Returning once more to Michell’s (2001) definition of cognition, ‘organismic sensitivity to the propositional structure and content of environmental situations’, it can now be seen how cognitive relations differ from other types of relations, even other types of sensitivity relations. Non-cognitive sensitivities are directed at less of the environmental features than are required for cognition to occur. For example, Michell (2001) considers an organism’s sensitivity to the sun’s being hot, as evidenced by its getting sunburnt. This sensitivity, he argues, is a non-cognitive one, because, “The sunburn is sensitivity to one and only one feature of this situation (the heat); it is not sensitivity to the situation’s entire propositional structure and content”. In order for perception of the ‘sun’s being hot’ situation to occur, an organism must be sensitive to the entire ensemble of features enumerated above (i.e., to space, time and the categories) as they occur within this situation. That is to say, it must be able to locate the subject, the sun, in time and space; to recognise the relevant property of the sun, its heat; to realise that there is just one subject involved; to know that the subject is predicated by the relevant

26 In considering that in order for cognition to occur, an organism must be sensitive to the spatial, temporal and categorical features of an environmental situation, it should be understood that these features are not firstly cognised individually, and thereafter actively assembled into a proposition, rather that, in cognising, the organism is sensitive to the propositionality of the situation as a whole.
property, that the sun is hot, as opposed to the subject and predicate being somehow disconnected; and to be suitably sensitive to the other categorial features of the situation.

Furthermore, this refined characterisation of cognition is illuminating with respect to the question of who, or what, the candidates for being cognitive referents are: namely, those systems that are sensitive to the propositional structure and content of environmental situations. This means that, although intuitively it seems that only a living organism could cognise, in fact it is logically feasible that machines could do the same job. Eschewing, for the moment, the possibilities of the future, it appears that all present-day cognisers are, in fact, organic, indeed, animalistic. Of course, this judgement is a difficult one to make because it is not yet certain what, exactly, the signs of cognition are. Of those who do appear to comply with the Direct Realist definition of cognition, language users, as Davidson argues (albeit for different reasons), are the most obvious, if only because the learning of language requires cognitive sensitivity to certain features of the world (symbols, things symbolised, and the conventions of symbolism). But it seems foolhardy to suggest that language usage is the only sign of cognition just because it is the one with which we are the most familiar. Additional signs of cognition are discernible in the alterations that occur in animals’ behaviour, the improvement in seeking and capturing prey, for example, that young animals undergo as they mature. Such improvement, it appears, could only take place if animals were capable of learning that ‘this creature is edible’, and other related facts.

On the other hand, there are some reactions that are made by plants and physiologically simple animals to environmental conditions, such as the presence of light, that need not be evidence for cognition, requiring, as they do, sensitivity to less than the situation’s propositional structure. Thus, for example, one need not look to cognition as a causal factor in order to explain why some plants react differentially to the presence and absence of light, as evidenced by their growing in the direction of its source (phototropism). Rather, the weakening of the cellulose that is contained within the cell walls on the side of the plant that is not exposed to light allows those cells to stretch and elongate, with the plant thereby bending towards the light (Ringo, 2001). What is required for this adaptive response is not, it seems, sensitivity to the propositionality of
the situation ‘the sun is in the north’, but simply the activation of relevant chemicals and enzymes.

In summary, the Direct Realist conception of cognition is that it is a non-symmetrical, binary relational situation that occurs when referents (thought at present to include only higher animals) become sensitive to the propositional structure and content of environmental situations. As such, cognition occurs externally to the referent of cognition, the organism, or, more specifically, the organism’s brain states, because the situation that forms the relatum of the relational situation is always external to the referent, and at no point becomes a quality or state of it. ‘Direct Realism’, it can be seen, is so named because it maintains that cognition is a direct connection between organisms and their environments, that is, because it denies the need for cognition to be mediated by internal ‘mental representations’, and also because it maintains that the objects of cognition, environmental situations, exist independently of being known, that they are real, objective existences.

Any Direct Realist theory is committed not only to the view that cognition is relational, but also to the view that such relations are subject to determinism, and are therefore incompatible with self- or goal-directed theories of behaviour, all of which it will deem to rest on mistaken premises, and thus to be false. The task of providing an account of mental causation from a Direct Realist perspective is therefore one of demonstrating that cognition forms part of the web of efficient causation. Maze’s (1983) attempt to do just this will be described in the following section.

4.4.3 Maze and mental causation
The deterministic premise defended and utilised by Maze in his investigation of human behaviour is stated negatively as follows, “… to say that organismic behaviour is guided by cognition does not entail that it is not caused in every detail, nor that it exhibits some privileged kind of self-generated ‘causality’ not enjoyed by the ordinary objects of nature” (p.8).

Having argued that mental causation is, like all causation, efficient, Maze (1983) seeks to discover what the variables are that are causally relevant to behaviour. The task that he sets himself is to “… give an account of motives as intrinsically characterised states
of the organism … in such a way as to make it intelligible that those states can be *caused* to occur, and that being in them the organism has its behaviour *caused in it* by its perception of its immediate surroundings…” (p.41). Maze comments further that the accomplishment of this task, like any scientific causal enquiry, requires one to “… discover both the relevant intrinsic properties of the organism and the relevant properties of something that arrived at or impinged on it at or just before t and caused the change that we are calling behaviour $B$” (p.47). In other words, Maze believes that the answer to the question ‘what caused behaviour $B$?’ requires the discovery of both the causal field (the relevant intrinsic properties of the organism) and the causal event (the something that arrived at, or impinged on, the organism). In this vein he suggests that,

For the psychologist the relevant properties of the organism will mainly be its motivational or drive state, which as we have seen must be characterised intrinsically, and its beliefs about the likely consequences of various behaviours in the situation it is in. What arrives as the external cause will typically, in fact invariably, be stimulation from the objects around the organism impinging on its sense organs, causing its perception of them, which will in turn recruit its acquired beliefs about those objects and what can be done with them relevant to relieving its drive state (p.47).

A critical feature of Maze’s (1983) account of behaviour causation that must be taken note of is that the organism’s acquired beliefs are understood to be efficacious by virtue of their being physiologically realised as state variables, that is, *brain states* that are intrinsic to the organism. He writes that, “For a person’s beliefs, conscious and unconscious, to play a part in the determination of behaviour, the brain states underlying them must somehow become involved in determining the course of the nervous impulses that will eventually give rise to motor processes” (p.87). Nevertheless, Maze has taken great care, as seen in Section 4.2.2.1, to establish the fact that brain states do *not* contain their cognitive relations as intrinsic features, and therefore sees fit to repeat that, “… those brain states should not be called the person’s *beliefs*, since the beliefs are the relational properties subserved by the brain states, their other terms being the facts believed in…” (p.88). What we are left with, then, is the contention that, insofar as cognition is causal, “… it is non-intentional *brain* processes which are at work” (p.5).
This conclusion gives rise to the puzzle that when cognitive relational situations cause behaviour, they do so not as cognitive relational situations in their entirety, but by means of just one of their terms, the brain states. Maze attempts to solve this puzzle by appeal to the notion of brain states’ possession of ‘relational properties’.

The concept of a ‘relational property’ is explored by Milton Fisk (1972), who aims to defend the thesis that relatedness does not require the existence of relations as entities that exist independently of relational situations. According to Fisk, relatedness is to be understood in terms of relational properties and their foundations. He illustrates what is meant by a ‘relational property’ as follows, “If six is greater than five, then six has the relational property greater than five, which is to be distinguished from the relation greater than, were there such an entity” (p.143). Fisk then turns to the foundations of relational properties, “To each relational property there correspond several foundations. To six’s relational property greater than five there correspond the magnitudes of six and five. One foundation alone is insufficient for a relational property; it is six having its magnitude and five having its magnitude that is the basis for six having such a relational property. The same pair of foundations in the relata will also be the basis for a relational property in five, the property smaller than six” (p.143).

In consistency with the claim that in order for a term to have a relational property, there must be more than one foundation (in the above example, in order for six to have the relational property greater than five, there must exist two foundations, the magnitudes six and five) Fisk (1972) insists that only the actual term in question can have a particular relational property. Thus, while it may seem that seven shares with six the relational property greater than five, Fisk argues that this is not so. This is because pairs of relational properties co-imply each other (from above, six’s relational property greater than five implied five’s relational property less than six). Taking $f$ to indicate a relational property, Fisk writes that “…$f$ here is some fixed relational property such as less than ten, and as such it is correlated with, that is, it coimplies, another fixed relational property, say greater than eight. Thus only the number eight has this particular relational property less than ten. The relational property less than ten that the number nine has is not a similar one since it is correlated with the dissimilar property greater than nine. Moreover, the foundation $F$ here is some fixed foundation, say the magnitude of eight, and clearly not every number has this magnitude” (p.145).
Relational properties are, it seems, to be identified by pairs (or greater numbers) of foundations (such as magnitudes) that are found to characterise separately existing entities, and, importantly, the co-implications existing between these, “Relatedness does not consist in the foundations just by themselves but in the correlation of the relational properties based on the foundations” (p.144). Accordingly, a single entity cannot possess relational properties independently of other entities in the way that it does with its regular properties. Thus, to borrow again from Fisk (1972), “… when Smith is taller than Jones, Smith’s having the relational property taller than Jones will coimply Jones’ having the relational property shorter than Smith. But, where their respective heights are the foundations, Smith’s having the height he has will not imply, or be implied by, Jones’ having the height he has. Still, taken together, the foundations coimply the relational properties” (p.143). Again, Fisk warns that, although one might be tempted to associate many heights, not just that of Smith, with being taller than Jones, it should be remembered that it is a property of Smith’s that is in question, and that “… this property is correlated with shorter than Smith, and is thus not similar to Hayes’ property taller than Jones, since this is correlated with the dissimilar property shorter than Hayes” (p.144).

Returning to Maze’s theory, the use of relational properties appears to be aimed at allowing an efficient causation account of behaviour production that recognises at once the relational character of cognition and the need for intrinsic physical, rather than mental, properties, as the efficacious causal element. Maze (1983) writes that, “In allowing for the causal role of cognitive processes in a deterministic account of behaviour, they are to be conceived of as state variables (but in the form of relational properties, as I am about to argue) of the organism, produced by previous interactions with environmental objects, and, along with drive state and other organismic variables, determining the behavioural effects produced in the organism by subsequent environmental stimulation” (pp.75-76). By conceiving of cognitive relational situations as relational properties of organisms, Maze seems satisfied that he has demonstrated that cognitive relational situations are unmysteriously causally efficacious.

So what does it mean for an organism to possess cognitive relational properties, and how does it come to acquire them? According to Maze (1983), exposure to events in one’s environment through sensory stimulation results in the formation of a brain trace,
this causal relationship sufficing for a cognitive relationship between the organism and
the situation to which it is exposed. The following example illustrates Maze’s claim,

An event to which my senses are exposed (a tree falls nearby) causes me to
perceive it. That is a psychological description, but what has happened can also
be described in physical-physiological terms referring to vibrations of the air
affecting my hearing mechanisms and thus the auditory area of the brain. (That
whole complex is the hearing; the hearing is not something extra that occurs
after the sequence reaches the auditory cortical area.) This physical-
physiological event results in a lasting brain trace having a specific relation to
the tree’s having fallen; the relation between the brain trace and the external
event can be spoken of as my knowing the external event, or at least as my being
in a condition such that I can remember (though not necessarily consciously
remember) that event if it becomes relevant to anything I am doing, i.e. if I see
something else as related to it (p.94).

Thus, according to Maze (1983), the complex causal occurrence whereby air vibrations
affect the hearing mechanisms and auditory areas of the brain results in a brain trace
that is unique to this particular sequence. In so doing, it provides the organism with the
foundation (more specifically the relevant brain trace) by which it either knows, or has
the potential to know, that the tree fell. Thus, it is through possessing the brain state that
enables its knowledge of the tree-falling event that the organism is deemed to possess a
cognitive relational property. Nevertheless, Maze does not mean to suggest that because
organisms have cognitive relational properties, they are anything other than
physiologically composed, nor that their being sustained by physiological processes in
any way lessens the psychological character of cognitive relational situations. He writes
that, “The fact that it is neurophysiological processes which subserve or sustain the
organism’s cognitive relations does not negate the fact that the latter are of that
distinctive kind that we call mental, that is, are the registering of propositional
happenings” (p.94). It appears, then, that by introducing the notion that brain states have
cognitive relational properties, Maze means not to deny that they are entirely composed
of physiological properties, rather, to acknowledge that they are composed of
physiological properties and are the foundations for organisms to enter into cognitive
relational situations.
It was stated above that Maze (1983) believes that a thorough account of mental causation requires the identification of both the relevant intrinsic properties of the organism (the causal field) and the causal event (the thing that impinges upon the organism). It can be seen now that, insofar as the former is concerned, the primary drives are central because they integrate the organism’s cognitions and muscular activity, being both knowers and biological engines.

Regarding the causal event, Maze (1983) understands this to be, “… stimulation from the objects around the organism impinging on its sense organs, causing its perception of them, which will in turn recruit its acquired beliefs about those objects and what can be done with them to relieve its drive state” (p.47). In this we see an important aspect of Maze’s theory, that all behaviour is a response to an object that is presently perceived, that, “… everything I do is a response to some stimulus object in my immediate environment” (p.70).

Hence, Maze’s (1983) suggested mechanism of mental causation involves a confluence of factors including the organism’s state of drive activation, its possession of intrinsic brain states which, although comprised only of physiological properties, enable the organism to have cognitive relational properties, and present sensory stimulation. He summarises his argument as follows,

The position is that an experienced organism will be in possession of a fund of information as to the general nature and potentialities for action of most of the things with which it comes into contact, and in consequence the afferent perceptual processes arising from things perceived here and now will, if the present objects are perceived as having properties relevant to what the organism is doing, be shuttled off into motor channels other than those they would have taken if the neural ‘information’ traces in the intermediary brain tracts had been other than they are; that is, if the organism had in the past acquired different information (pp.92-93).

Thus, if Maze is correct, it is through perceiving present situations, and already knowing facts relevant to these situations, that the primary drives, when activated, are made to respond with movements that, for the most part, result in their own
deactivation. Importantly, because each phase in the causation of such consummatory actions is brought about by necessary and sufficient present and precedent conditions and events, the postulation of ‘final’ or teleological causes, or mysterious human capacities to act independently of efficient causes, is unwarranted. Nevertheless, there remain some concerns that, as a theory of mental causation, Maze’s account is unworkable because, firstly, it suggests that it is organisms’ possessions of non-intentional brain states, rather than their being in cognitive relational situations, that is causally relevant, and secondly, because the notion of ‘relational properties’ is problematic. These criticisms will be developed in Section 5.1, and a solution to them will be attempted in Section 5.2.
Section 5: Further Development of a Direct Realist Theory of Mental Causation

5.1 Maze, ‘relational properties’ and epiphenomenalism

Evident within Maze’s (1983) writing is an assumption that appears to have been of considerable influence in restricting the kinds of things that he is willing to consider might be involved in efficient causal processes, namely, the assumption that it is an object’s possession of *intrinsic* states, properties and processes, and not its standing in relations to other things, that makes it causally efficacious. Now obviously if, as Maze suggests, phenomena that are relationally defined are *not* the kinds of variables that make plausible candidates for behaviour-causation, then cognition, which is relational in character, cannot be causal in this context. Accordingly, Maze believes that in order to be viable, a theory of psychological determinism must identify the intrinsic organismic variables that provide the mechanisms by which desires and cognitions work. He writes that, “If one is to make good the rescue of ‘belief’, ‘desire’, and ‘intention’ from the merely dispositional status to which they had been consigned by philosophical behaviourism, then one must give some indication, in however general or schematic terms, of what kind of thing the substantive basis of the behavioural dispositions might be – that is, what kind of intrinsic property, process, or state these internal entities ‘belief’, ‘intention’, and ‘desire’ might turn out to be” (p.4).

As has already been indicated, Maze (1983) understands the traditional role of desires, that of motivating behaviour, to be best accounted for in terms of the physiological structures and processes of the primary drives. Similarly, he argues that cognition’s causal role, that of guiding behaviour in accordance with the facts of organisms’ environments, is realised through the occurrence of various brain states and processes. Maze’s claim that the brain state variables that are relevant to behaviour-causation are to be conceived of as ‘relational properties’ appears to be aimed at providing an account of mental causation that utilises the facts of physiology but that does not imply a form of physiological reductionism. Were the concept of relational properties a sound one, then Maze could defend the view that an organism’s cognitions are in some sense internal state variables, and as such he could argue that they are involved unproblematically in the production of behaviour. But, as I aim to show in what
follows, the concept of relational properties is not sound, and therefore Maze’s ‘internalist’ assumption is incompatible with cognition causing behaviour. In other words, it is the aim of the present section to demonstrate that Maze’s internalist assumption, whereby it is through possessing certain *intrinsinc* variables that an organism’s cognitions are causally efficacious, results in a theory of behaviour-causation in which cognitions are rendered epiphenomenal.

The problem with the suggestion that objects possess relational properties is that it confuses a situation in which at least two terms interact in some manner with a property or quality of one of those terms. Concerning cognition in particular, to suggest that beliefs and so forth are relational properties of organisms is to suggest that the cognitive relational situations into which organisms enter with environmental situations are brain states, that is, characteristics of the organism alone. This in turn paves the way for a representational account of cognition, one in which the organism’s brain states are theorised to refer to, or represent, the environmental situations that are known, rather than simply entering into a cognitive relational situation with them. Maze (1983) is well aware that the attempt to depict cognitive relational situations as properties of organisms invites interpretation of this sort, as evidenced by the following, “The problem in talking of cognitions or beliefs as state variables is to avoid giving the impression that one is embracing a representationist theory of knowledge, in which the state variables are internal representational entities” (p.76). What Maze seems not to recognise is that the postulation of relational properties, whether of a cognitive or alternative character, involves a logical error, and therefore that such things cannot exist.

The logical problem with treating an organism’s cognitive relations to environmental situations as intrinsic or state properties of that organism is that one engages in a category mistake whereby a situation that extends beyond the confines of the organism’s being is treated as if it were a component of it, that is, a characteristic that contributes to the organism’s identity. However, this practice ignores a vital distinction between properties and relations, the fact that whereas the organism’s intrinsic states or properties do not imply the existence of anything beyond themselves, its standing in
relations does. 27 Accordingly, the requirement that any relational situation involve at least two terms that interact in some specific manner (in cognition, the knower, the environmental situation that is known, and the ‘knowing that’ manner of the interaction), establishes that no such situation could be intrinsic to just one of the terms, because no specification of it could be made without reference to both the other term and the character of the relation itself. The error involved in postulating that an object possesses relational properties is therefore a matter of confusing situations that extend beyond that object, and necessarily involve other objects, with internal states that belong exclusively to that object itself.

To avoid this confusion, what requires emphasis is the fact that, while it will always be the case that in order to participate in a relational situation an object will need to possess certain properties, these latter are, in Fisk’s (1972) terminology, simply the ‘foundations’ for the relational situations, and are in no sense relational themselves. In other words, while the possession of foundational properties, such as brain states and magnitudes, enables objects to participate in relational situations with other objects, no object ever contains, in addition to these properties, relational properties such as ‘knowing that $p$’ or ‘being greater than five’. This, of course, is of immediate relevance to Maze’s (1983) theory because it indicates the difference between acknowledging that certain organismic states and processes are involved in cognising the world, and the situation whereby knowing facts concerning one’s environment influences one’s reactions to that environment. I shall elaborate.

As was detailed in Section 4.4.2.1, Maze (1983) argues that the referents of cognition are certain organisms, and that organisms that cognise are able to do so because they possess the requisite intrinsic properties, namely brain states. It was also remarked that these brain states should not be confused with the cognitive relational situations themselves because they (the brain states) are but one term of those relational situations. In Maze’s words, “By cognitive processes I mean believing, perceiving, knowing that, being conscious of, remembering, and so on; these can coherently be thought of only as relations into which the organism enters, or more specifically as relations between certain specific kinds of bodily process (primarily, brain-plus-sense-organ processes)

27 As Anderson (1944) claims, “We regularly make a distinction between qualities and relations in that a relation holds between two things, whereas a quality does not raise the question of a second thing.”
and things external to those processes” (p.83). What is most important about this characterisation of cognition for the present discussion is the aforementioned distinction that exists, and is recognised by Maze, between brain states and cognitions, “… those brain states should not be called the person’s beliefs, since the beliefs are the relational properties subserved by the brain states, their other terms being the facts believed in…” (p.88).

It is difficult to comprehend then, given the above, why Maze, who is perfectly aware that cognitions must be relational situations that are external to brain states, would continue to maintain that it is only by discovering their substantive bases in intrinsic variables that an organism’s cognitions can be rescued from a mere dispositional status. Nevertheless, the postulation of brain states and processes by which cognitions are efficacious is what Maze’s internalist assumption demands, and cognitions are accordingly (but erroneously) likened to organismic properties. The end result is that, rather than elucidate the role that thought plays in the production of behaviour, Maze’s theory relates instead only the part played by brain states, their neurophysiological properties and processes.

This focus upon the role played by brain states, at the expense of that played by cognitive relational situations, can be discerned in Maze’s (1983, p.94) example wherein the falling tree causes his perception of it (the quotation can be found in Section 4.4.3). According to Maze, the perception of the environmental event causes the formation of the brain state, and the brain state that is thereby formed has at least two potential effects. Firstly, it can become involved in the determination of neural and muscular events. Secondly, it provides the potential for the organism to know that the tree fell. This second effect, the enabling of the organism’s knowledge of the tree-falling event, is seemingly superfluous to the brain state’s role in determining how the organism behaves because it is simply an effect, in addition to the neural and muscular ones, of the brain state’s coming into existence. If this reading of Maze is correct, then because they are both effects of the same cause (the formation of the brain state), neither the organism’s knowledge, nor its bodily movements, is causally necessary for the other, or, in other words, what the organism knows is causally irrelevant to what it does.
This brings us to the conclusion that, in Maze’s (1983) theory, whilst brain states are central to the determination of behaviour, or, as Maze expresses it, whilst “… it is non-intentional brain processes which are at work” (p.5), cognitions are tangential, the most generous possible reading allowing only that they have some sort of connection with these efficacious brain states because they are subserved by the same, “For a person’s beliefs, conscious and unconscious, to play a part in the determination of behaviour, the brain states underlying them must somehow become involved in determining the course of the nervous impulses that will eventually give rise to motor processes” (p.87). But even this reading shows that, as far as behaviour-causation is concerned, an organism’s participation in cognitive relational situations (knowing facts about the world), whilst dependent upon the same physiological states and processes that cause behaviour, is itself causally epiphenomenal.

This implication, that cognitive relational situations do not have a causal role in Maze’s account of behaviour-causation, is recognised by Michell (1988) who writes that, “If relational situations cannot be causes, then cognition does not have a genuine role in the production of behaviour” (p.241). All that could be left for cognition to do, under these circumstances, is to “… be useful in predicting behaviour because it is correlated with the genuine cause (viz. neural states of the brain) and, given present technical limitations, can sometimes be known more reliably than the relevant neural states” (p.241).

Now, if indeed every cognitive relational situation does have a brain state peculiar to it, then perhaps cognition could be a discernible indicator of the occurrence of the causally relevant brain states, and in being so discernible be in some sense explanatory in a way that the mere occurrence of brain states is not. This, at least, is the difference between cognitions and brain states that Maze appears to be drawing attention to by the following. “So, then, when I say that a person’s beliefs constitute a set of state variables that in part determine his or her behaviour, I mean that the person has a large collection of brain states specific to those beliefs, but of these brain states we know nothing directly and so their existence is of no practical use in explaining his or her behaviour” (p.88).
But having the entirety of its worldly impact concentrated in this ancillary role seems to be a far cry from what Maze has envisaged cognition as doing, that is, as partaking in causal relations both as effect (that is, through being determined itself) and cause (that is, through determining behaviour). Indeed, Maze’s commitment to cognition as a causal variable is evident in the difficult task that he has undertaken with regard to explaining how cognising organisms interact with the propositional structure and contents of their environments, that is, to how the behaviour of organisms is guided by their perception of the facts around them. It is this commitment, I believe, together with his internalist assumption that it is only by its possession of certain brain properties that an organism can be caused to behave, that has driven Maze’s conception of cognitive relational situations as ‘relational property’ state variables.

However, as has been argued, the concept of relational properties is unsound, which leaves us with the conclusion that, because there is nothing psychological characterising a brain state, or, as Maze expresses it, because “Any such specific brain state would simply be whatever it physiologically is…” (p.86), no account of mental causation can be given in terms of intrinsic brain states and processes. To suggest that an account of mental causation can be given in terms of what the brain is and does is only to ignore what distinguishes mental events from processes occurring within an organism’s brain, this being that they are relational situations involving environmental facts and as such occur, in part, externally to brain states.

In conclusion, when it is an account of how relational situations such as cognitions are causally relevant that is desired, the appeal to relational properties obscures the need to discover not what intrinsic organismic variables are at work in mental causation, but what relational situations, external to organisms’ brains, are functioning so. The problematic nature of the appeal to relational properties reveals itself in Maze’s (1983) theory through the allowance of unwanted representationist interpretations and, what is more, mental epiphenomenalism because brain states, rather than cognitive relational situations, are seen to do all of the causal work. Accordingly, it is to be argued in the following section that the revision required for Maze’s theory to become a viable approach to mental causation involves the introduction of cognitive relational situations as causes.
5.2 An alternative Direct Realist account of mental causation

One of the most challenging tasks that an externalist theory of cognition such as Direct Realism faces is to provide a convincing account of how coming to know something about the world causes an organism to behave in one way rather than another. If such a theory is to be found viable, it must avoid the temptation to redefine cognition in an internalist fashion whereby thought is treated as a property of the cognising organism, and it must also demonstrate that relational situations, as well as state variables, can be causally efficacious.

I refer here to the practice of redefining cognition as a property of the organism as a ‘temptation’ because this option is the one that is most often retreated to by ‘externalists’, at the expense of a genuine theory of mental causation. Having accepted that cognition is external to the cognising organism’s body, I propose to pursue a different option, one that examines more closely some of the issues concerning the character of causation, most particularly as concerns the issue of whether relational situations can, or cannot, be causal. Preliminary to this, however, there is a need for clarification concerning why externalism has appeared to so many theorists to be incompatible with behaviour-causation. Accordingly, I shall consider firstly the popular assumption that only ‘local’ variables are involved in causation (Section 5.2.1), secondly, the character and location of behaviour (Section 5.2.2), and finally, the prospects for cognitive relational situations being causally efficacious in behaviour-production (Section 5.2.3).

5.2.1 Why causation and externalism seem to be incompatible: the ‘locality assumption’

The seeming incompatibility of externalism and behaviour-causation is stated perspicuously by Colin McGinn (1989), whom I shall quote at length.

---

28 Most typically, theories that begin by stating that what is known exists outside of the knower end with the adoption of either a form of representationalism, whereby physiological brain states are thought to ‘refer to’ or ‘be about’ something external to them, and thus in some sense to incorporate, within the knower, what is known, or with a form of ‘strong supervenience’, which ultimately leads to psychophysical reductionism. At this point, not only do the theories cease to be genuinely externalist, they also invite the claim that it is the ‘syntax’ (i.e., physiological structures) and not the semantics (i.e., meanings) of brain states that is causally efficacious, thus defeating the purpose of demonstrating that content can be causal. For examples of theories that have sought a solution to the problem of externalism in representationalism and strong supervenience see Dretske (1993) and Kim (1982) respectively.
The argument proceeds from certain principles about causation (so let us call it ‘the causal argument’) – specifically about how mental causation must be seen to work. And its conclusion is that content and causality do not mix: content can play no (nonredundant) causal role in producing the effects of a given mental state, either behavioural or mental. In a word, content is causally epiphenomenal. The argument for this claim presupposes externalism, either weak or strong. According to externalism, contentful states are identified by reference to entities that lie outside the subject’s body (including, of course, his head): these extrinsic entities are the constituents of such states. Now this implies that contentful states are (as we might say) extrinsically relational: they essentially consist in relations to extrinsic nonmental items (objects, properties, etc.). They thus involve certain ‘correspondence relations’ between the subject (or his intrinsic states) and entities that exist beyond the confines of the subject… (pp.132-133).

And now the causal objection to citing such contents in psychological explanations is just this: if such explanations purport to be causal, then these correspondence relations cannot themselves be implicated in the causal transaction being reported. This is because what happens at the causal nexus is local, proximate and intrinsic; the features of the cause that lead to the effect must be right there where the causal interaction takes place. Causation is the same with brains and minds as it is with billiard balls. Their effects depend upon local properties of these entities. The causal powers of a state or property must be intrinsically grounded; they cannot depend essentially upon relations to what exists quite elsewhere. The question then is how correspondence relations to what exists outside of the subject’s body can be causally relevant to changes initiated from within the body. To put it less tendentiously, how can contentful states have their effects in virtue of the extrinsic relations that constitute them? Surely if the external relata of these relations were not to exist, this would not necessarily alter the causal power of states of the subject to bring about the same bodily movements. At best, therefore, the extrinsic relations definitive of content are causally redundant, thus inviting an application of Occam’s razor. The causal mechanisms do not incorporate these relations, and causal explanation consists in describing (however crudely) the operation of such mechanisms (pp.133-134).
McGinn’s (1989) ‘causal argument’ captures the intuition which I believe to be behind much of the resistance to a genuinely externalist theory of cognition, the intuition that causes must be spatially and temporally contiguous with their effects. Whether or not causal contiguity is genuinely incompatible with externalist theories of cognition (an issue to which I shall return in Section 5.2.3), it is nevertheless apparent in the writing of many that it is assumed to be so. This is because externalism maintains that cognitions are relational situations which are in part comprised of environmental situations that are spatially and temporally distant from the cogniser, and, accordingly, that part of what causes an organism to behave is located at some distance from it. As a consequence, externalism is deemed to be suspicious because it is suggestive of ‘action at a distance’ or some other transgression against the contiguity intuition. In Frank Jackson’s (1996) words, “Externalism about content violates the here-and-now intuition” (p.401). The obvious alternative to accepting this ‘violation’ has seemed to be to “… show that each intentional state needs some intrinsic feature or features that serve as the mechanism for its interaction with other states, and ultimately with actions. These intrinsic features will be the local causal surrogates for the contents of the states” (Crane, 1992, p.197), in other words, to abandon the externalist position.

However, abandoning the externalist position means abandoning the basic premise of Direct Realism, that what is known in cognition is an environmental situation that exists independently of the knower, and, accordingly, when it is a Direct Realist theory of mental causation that is desired, options other than rejecting externalism must be explored. One of these is to examine more closely the character of causal contiguity and in so doing to discover if it is genuinely incompatible with relational situations acting as causes, in other words, to address the assumption that only ‘local’ or ‘intrinsic’ variables are involved in causation.

One useful way of approaching the question of whether only intrinsic variables are involved in causation is to consider the constitution of causal mechanisms. According to Machamer, Darden and Craver (2000), mechanisms are “… entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (p.3). Furthermore, descriptions of mechanisms are made intelligible by their demonstrations of ‘productive continuity’,
Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to termination conditions. Productive continuities are what make the connections between stages intelligible. If a mechanism is represented schematically by $A \rightarrow B \rightarrow C$, then the continuity lies in the arrows and their explication is in terms of the activities that the arrows represent. A missing arrow, namely, the inability to specify an activity, leaves an explanatory gap in the productive continuity of the mechanism (Machamer et al., 2000, p.3).

Now, presumably, Machamer et al. (2000) suggest that productive continuity is necessary for the intelligibility of descriptions of mechanisms because productive continuities are features of mechanisms as they operate in the world. Granting the need for productive continuity in a causal mechanism, one is brought immediately to the question, is such continuity to be equated with the spatio-temporal contiguity of the successive stages of the mechanism, that is, with spatio-temporal contiguity between causes and effects? This question is a very difficult one to answer. Mackie (1980) suggests that, “While we are happiest about contiguous cause-effect relations, and find ‘action at a distance’ over either a spatial or a temporal gap puzzling, we do not rule it out. Our ordinary concept of causation does not absolutely require contiguity; it is not part of our idea of causation in a way that would make ‘$C$ caused $E$ over a spatial, or temporal, or both spatial and temporal, gap, without intermediate links’ a contradiction in terms” (p.19). Be this as it may, Mackie states much later in his book that, “…contiguity, that is, a merely spatio-temporal continuity of process, may have a place in causation as it is in the objects…” (p.223), in other words, while it is not essential to ordinary concepts of causation, contiguity may nevertheless, as a matter of fact, feature universally in causal processes.

Accepting, at least provisionally, the need for causes and effects to be spatially and temporally contiguous, we come now to the question of whether only the intrinsic properties of objects participating in causal relations are of causal relevance, or whether relational situations that extend beyond these objects are eligible causal candidates too. The appeal of the intrinsic properties is that they seem, at first glance, to be genuinely

29 Consider, for example, the gravitational force that is exerted between two bodies. Gravitational force is a product of the masses of the two bodies divided by the square of the distance between them. Here, it appears that the mass of one body is causally relevant to the acceleration of the other, at a distance.
local, right where the action is, but of course properties, like objects, are extended in
time and space, and therefore the suggestion that they are truly ‘local’ is misleading if
by this it is meant that they occupy something less than a spatio-temporal region
(something, for example, like an unextended point in space and time).

A brunette’s property ‘brown’, for example, is extended throughout the spatial region
occupied by her hair, and the temporal region extending from when the brown hair
appears to when it falls out, or goes grey. Thus, while it is correct to say that an object’s
properties are contained in a sense that its relations are not, in the sense, that it is to say,
that they do not extend beyond the individual objects, it is incorrect to say that the
properties are more local, spatio-temporally speaking, than its relations, because they
need not be. It is perfectly feasible, in other words, that a property such as brownness
could occupy a greater spatio-temporal region than does a relational situation, as occurs,
for example, when a woman who is a brunette for fifty years has a hair on the crown of
her head which is longer than a hair at the nape of her neck, for just ten years. As
demonstrated here, a relational situation can occupy shorter temporal durations, and
lesser spatial dimensions, than does a property. Of course, any combination of greater
and lesser spatial and temporal extents may occur for the spatio-temporal location of
both the property and the relational situation, the point being that there is no necessity
that an object’s possession of intrinsic properties be any more ‘local’ than its
participation in relational situations.

As has been implicit in the above, an object’s being propertied is a situation, just as is
any relational connection it might have to another object. Once this is realised, it
becomes evident that it is in no sense possible to distinguish situations whereby ‘objects
possess properties’ from situations whereby ‘objects stand in relations’ in terms of
localness, because both are spatially and temporally extended units. Accordingly, if, as
was argued in Section 4.3.1, it is the case that the referents of causation are events, then
it must be recognised that even when properties are causally efficacious, they are so not
as isolated properties per se, but as components of complex situations that are involved
in the event of impinging upon a causal field. This being the case, rather than deny
relational situations eligibility as causes because of their spatio-temporal extensions,
one must recognise that all causes are extended in this manner, because all events
involve spatio-temporally extended situations.
What implications does the foregoing have for the productive continuity and spatio-temporal contiguity of causation? One implication might be that, in granting the need for spatio-temporal contiguity between causes and effects, one must recognise simultaneously that only parts of these situations ever actually come into contact. Perhaps this is not so difficult to come to terms with when we consider everyday examples of causation, such as when a thrown brick smashes a window. Here, just one surface of the brick actually comes into contact with one part of the window, but one would not want to conclude from this that only the properties located on that surface of the brick were responsible for the breaking of the window. Rather, an entire complex situation involving the brick’s having a certain mass, inflexibility, rate of movement and, indeed, relations to the window (such as impacting upon it) is the cause, and, of course, it would be false to claim that this situation is located on the surface of the brick that contacts the window.

Furthermore, recognition that spatially and temporally extended relational situations are causes might be the key to demonstrating that apparent cases of ‘action at a distance’, such as those in which spatially distant bodies exert gravitational forces on each other (see f.n.29), are better understood as ‘action across a distance’, because they involve relational situations such as ‘body $a$ is distance $x$ from body $b$’ as causal factors. Rather, then, than think that it is simply the masses of the objects that are causally relevant to their mutual attraction, we see that it is a situation involving their respective masses as well as their being a certain distance from each other, that accounts for this effect.

In conclusion, the spatio-temporal contiguity involved in the productive continuity of causation is realised by events wherein complex situations are introduced into the causal fields within which they produce their effects. Because all situations are spatially and temporally extended, it is impossible to differentiate those that are characterised by an object’s being propertied in some way from those that involve two or more objects’ being related in some way in respect of the former being more ‘local’ than the latter, or, indeed, the latter being more ‘local’ than the former. Accordingly, it is an error to conclude, $a$ priori, that relational situations are incapable of acting as causes on the grounds that they differ unfavourably from situations involving states and properties in terms of their ‘localness’. What can be concluded is that the ‘locality assumption’ of
5.2.2 The character and location of behaviour

While much of the mental causation literature is concerned with discovering the kinds of variables involved in behaviour-causation, very little attention is directed towards the difficult problem of defining what, exactly, behaviour is and where, precisely, it occurs. Of course, everyone has at least some notion of what it is to behave badly or courageously, treacherously or with honour, yet when it comes to defining, in the most general terms, what is common to all behavioural acts, the task is revealed at once to be of great complexity. Consider, for example, Skinner’s (1938) definition of behaviour,

> By behavior, then, I mean simply the movement of an organism or of its parts in a frame of reference provided by the organism itself or by various external objects or fields of force. It is convenient to speak of this as the action of the organism upon the outside world, and it is often desirable to deal with an effect rather than with the movement itself, as in the case of the production of sounds (p.6).

In this passage, the character of behaviour is taken to be discovered in, firstly, *bodily movements* within an environmental context, and then, secondly, the *effects* of these movements upon the environment, rather than the movements themselves. Already puzzles have arisen as to what behaviour might consist in because, although the two suggestions are somewhat at odds with each other (as indicated by the ‘rather’ in ‘rather than the movement itself’), neither is dispensable.

Considering, firstly, the second definition, it is clear that, in practice, the identification of behaviour requires consideration of the kinds of environmental effects that are produced because these have conceptual unity in a way that the bodily movements need not. For example, Skinner (1935) sets up the problem of the numerous *unique* combinations of bodily movements that result in the *same* operant response as follows, “The number of distinguishable acts on the part of the rat which will give the required movement of the lever is indefinite and very large” (p.351), and locates the solution to this problem in the fact that, “Except for certain rare cases they constitute a class, which
is sufficiently well-defined by the phrase “pressing the lever”" (p.351). In other words, it is the sameness of the environmental effect, not the bodily movements, that accounts for these acts being behaviours of the same type, lever pressing. However, this does not license one to define behaviour independently of bodily movements, because obviously if the environmental effect were to occur in some way other than through the movement of the organism, if, for example, the lever were depressed by the experimenter in the rat’s presence, then it could not be said that the rat was engaging in ‘lever-pressing behaviour’, despite the fact that the same type of environmental effect came to pass. Thus, although it may well be the environmental effect that is responsible for how behaviour is classified, as this kind of act or that, some input from the organism is always essential if there is to be any behaviour at all.

This brings us to the role of bodily movement in behaviour. Clearly not all bodily movements are behaviours (consider reflexes and metabolic processes), and possibly not all behaviour requires bodily movement (think of a stalked animal ‘playing dead’). Nevertheless, most behaviour involves bodily movement, and it is clear from the way in which the terms are used interchangeably in the mental causation literature that bodily movement and behaviour are generally considered to be equatable with each other. While this focus upon bodily movements in the explanation of behaviour is understandable, it is also problematic because, firstly, it is too coarse-grained to differentiate between what an organism does and what happens to it, and, secondly, because it locates behaviour in too narrow a compass. These problems will be considered in turn.

Dretske (1988) has directed considerable attention to the distinction between what happens to an organism (or machine for that matter), and what an organism does. For example, he writes that, “When a rat moves its paw, that is something the rat does, a piece of rat behavior. When I move its paw, the paw still moves, but the rat doesn’t move it. There is no rat behavior. Indeed, I could be moving the paw of a dead rat, and dead rats do not behave” (p.1). To extend this example, imagine that in both instances, the one in which the rat moves its paw and the one in which Dretske moves the rat’s paw, the paw ends up on top of a lever, the latter being duly depressed. Now, both events seem to accord with Skinner’s (1938) definition of behaviour. They both involve the relocation of a part of the organism with reference to the objects about it, and they
both result in a certain type of environmental effect, the depression of the lever. But, of course, no-one would be inclined to say that these two instances of paw relocation and lever depression each constitute ‘lever-pressing behaviour’. Rather, what must be concluded is that, even if behaviour does require the occurrence of bodily movements and environmental effects, there is something more to it, something that distinguishes the ‘rat moving its paw and thereby pressing the lever’ sequence from the similar one initiated by Dretske. This way of stating the problem points to a possible solution, one that Dretske has developed at some length.

The difference, Dretske (1988) argues, between a sequence whereby a rat moves its paw and a sequence whereby a rat’s paw is moved by someone or something else is that the first, but not the second, involves a cause that is internal to the organism. Furthermore, this distinction, involving the changes that are initiated from within organisms as opposed to those that are initiated from without, can be used, he argues, to characterise behaviour in general, “If we have a well-defined ordinary notion of behavior – and, aside from the vague contrast between things we do and things that happen to us, I am not sure we do – it is, with a few refinements, equivalent to internally produced movement or change” (p.3). As Dretske hints in this early definition, and argues for more explicitly in the pages that follow it, behaviour is constituted by environmental effects, bodily movements and internal organismic causes, and therefore cannot be equated with any of these to the exclusion of the others. In his words, “… the behavior is being identified not with the internal cause (C) and not with the effect – proximal (M) or remote (N) – but with a temporally more extended process: the one thing’s causing the other” (pp.21-22). This ‘behaviour as causal process’ definition, Dretske maintains, captures the distinction between the products of behaviour (movement (M) and environmental effects (N)) and the behaviour itself, and also accounts for the fact that behaviour can occur over great distances and vast time spans.

Dretske (1988) devotes much attention to the important distinction between what an internal organismic cause (C) causes (i.e., bodily movements (M) and more remote environmental effects (N)) and the process by which (C) causes (M) and (N). For example, he argues that, “It would be an elementary confusion to identify, say, a rat’s paw which was moving (= a moving paw) with the paw’s movement. The first is an object, a paw; the second is an event, a movement. It is the same confusion, though not
at so elementary a level, to confuse movements which are brought about by internal events with their being brought about by these events. The former is an event, a movement, something that happens to (say) a paw. The second, I shall argue, is a piece of behavior, possibly an action, something the rat *does*” (p.15).

The distinction that Dretske (1988) is highlighting here is between a product (paw movement) and a process (the moving of a paw). The process, he argues, has as a constituent the product, and, of course, the process is not complete until the product has been produced. Nevertheless, just because bodily movements and environmental effects are necessary constituents of behavioural processes, it would be mistaken to locate behaviour in these. Likewise, even though it might be tempting to locate behaviour where the internal cause occurs, particularly when the effects of this cause are spatially or temporally remote from the organism, this would be to ignore the endpoint of the process, and thus to mislocate it once more. In order to avoid neglecting any part of the behavioural process, either the organism’s contribution or the effects that are produced, Dretske concludes that, “Behavior, to be sure, requires some internal C to produce M, but that fact doesn’t require us to identify behavior with either the M (which is caused) or the C (which causes it). One can, as we have already done, identify behavior with a process – C’s causing M – that begins with C and ends with M” (p.17).

Furthermore, by conceiving of behaviour as a causal process, Dretske (1988) is enabled to address the fact that behaviour is extended in time and space, and, in so doing, to provide answers to potentially puzzling questions concerning when, and where, behaviour occurs. Consider the following example, “Turning on the lights normally occurs in a twinkling. You really don’t have time to do anything else after you flip the switch and before the lights go on. But locate the bulb on Neptune and the switch on earth. Now you have time to go to bed after you flip the switch and before the lights go on. Are you turning on the lights while snoring in bed? What if you die before the lights go on. [sic] Are you turning the lights on after you die? Can dead people do things? But if you didn’t turn the lights on, who did?” (p.20). Of course, turning on the lights is always a spatially and temporally extended causal process, but Dretske’s example is useful for highlighting what may well appear to be a counterintuitive consequence of viewing behaviour as a process that begins with the organism and ends with an environmental effect, namely, the fact that behavioural processes are not necessarily
contained in the vicinity of the behaving organism, either temporally or spatially. Focusing largely upon behaviour’s temporal span (although elsewhere he draws similar conclusions concerning its extended spatial location), Dretske concludes that, “… behavior, like any protracted event, condition, or process, is no more precisely datable than its (temporal) extremities permit… The temporal coordinates we use to locate behavior must have a “thickness” commensurate with the temporal spread of the behavior being located” (pp.20-21).

By recognising that because bodily movement is a constituent of behaviour the one is not equatable with the other, one is enabled to recognise likewise that, even though it is the organism that behaves, behaviour itself need not be entirely located where the organism is. The limits of behaviour, Dretske (1988) suggests, are located, at one end, where the organism is and, at the other, where the relevant environmental effects are. He writes that, by viewing behaviour as a process, one makes “… behavior begin where it should begin (with those efferent activities that bring about bodily movement) and end where it should end (with those external events or conditions that the behavior requires for its occurrence)” (pp.17-18). The merit, I believe, of this position, is that it captures the fact that behaviour is located in a region the boundaries of which are dictated by the spatio-temporal locations of the organismic cause and the environmental effects, no matter how far asunder these may be. In other words, behaviour is not somehow confined within the behaving organism, but is spread out within an environmental system of which the organism forms a part. Nevertheless, the identification of behaviour with a process that begins within the organism is problematic because it ignores the crucial role played by cognition. I shall elaborate.

It has been shown that, according to Dretske (1988), the difference between what an organism does and what it endures resides in the former sequence, but not the latter, being initiated by an internal cause. This characterisation of behaviour as a process that begins with internal organismic causes, and ends with environmental effects is, he believes, in accordance with thought playing a causal role in behavioural acts, precisely because he understands thought to be constituted by internal mental representations. If indeed thought were located within the organism, and if behaviour were, as Dretske maintains, a process that begins inside the organism, then it would be most natural to conclude that there was neither paradox nor mystery attached to mental causation
because cognitions would be found to be internal organismic states, and these are precisely the kinds of things that, when activated, would differentiate behavioural from otherwise similar non-behavioural sequences.

Tidy as this solution may seem to be, it of course relies upon the acceptance of a theory of mental representation, and the consequent abandonment of the externalist position. This ‘solution’, however, is more problematic than is the conundrum that it was invoked to resolve. Not only does the retreat to internal mental representations that act as surrogate objects of thought bring with it a whole host of untenable suppositions, such a move avoids rather than answers the question being posed concerning the way in which external cognitive relational situations are causally efficacious in behaviour-production. It is, therefore, no solution as far as a Direct Realist is concerned.

Despite this, there is much that appeals in Dretske’s (1988) analysis of behaviour as a spatially and temporally extended causal process. Perhaps, then, what is required is not the rejection of his theory in its entirety, but a re-examination of the definition of behaviour with regard to the role played by cognition.

Such an approach has been undertaken by both Maze (1983) and Michell (1988) in their efforts to distinguish accidental from non-accidental effects of bodily movements. As has already been demonstrated, the classification of a behavioural act as being of this type or that requires the identification of the environmental effects that the organism’s movements produce. Skinner (1935), we saw, claimed that any number of distinguishable sequences of bodily movements might be classified by their common environmental effect, the depression of a lever, as instances of ‘lever-pressing’ behaviour. But, of course, there remains the concern that, not only do all of these sequences have as a common effect the depression of a lever, they might also have in common any number of other effects, such as the causing of a clicking sound in the lever apparatus, none of which are in any way useful for classifying what the organism is doing.

Seizing upon this, Maze (1983) has suggested that, “… to say what someone is doing is to say what that person thinks he or she is doing” (p.100), or, in other words, that the classification of a behavioural act requires the identification of one or more non-
accidental consequences of the organism’s bodily movements. In his review of Maze’s argument, Michell (1988) elaborates upon the connection between cognition and the non-accidental consequences of bodily movements as follows, “… A’s doing X is non-accidental whenever A’s belief that bodily movements of kind M in these circumstances will bring X about is one of the cognitions guiding A in producing movements M and consequently X. Otherwise, A’s doing X is accidental” (p.235). By this account, we see that in order for an environmental effect to be non-accidental, and therefore of use in classifying what an organism is doing, it must have been caused by a belief that the organism had concerning it. In this manner, we observe not only that cognition is indispensable for the causation of behaviour, but also that the externality of cognition is the key to organisms’ effective interactions with the environments in which they are placed.

Concerning this latter observation, some comment is required. It was remarked above (Section 5.1) that Maze’s (1983) focus upon the role of non-intentional brain states as the agents of cognition in behaviour-causation is inappropriate because these neither contain nor refer to the environmental situations which form the objects of beliefs, and, such being the case, that the project undertaken by Maze to explain how organisms’ behaviour is guided by the environmental features of their surroundings is left unfinished. As was suggested there, what is required for a viable theory of mental causation is not an account of what intrinsic brain states do, but an acknowledgement of the role played by external cognitive relational situations in the causation of behaviour.

The necessity of recognising the externality of cognition in the context of behaviour-causation lies in the fact that behaviour is a form of interaction with the environment, one that would not be possible if facts concerning the environment were unknown. Through knowing details of the environment in which it is placed, for example, through its perception of the location of objects, an organism is enabled to react appropriately and effectively to that environment, by which means it can, conditions being favourable, satisfy its needs. Of course, no such connection with the world surrounding it could occur within the organism itself, and, accordingly, it is only because cognition is a type of relation that occurs externally to the cognising organism that it can be of use in guiding behaviour, which is likewise directed upon the outside world.
For example, when prompted by thirst, a man’s movements might be directed by his knowledge that water is to be found in the kitchen such that he arrives in the kitchen, rather than in some other place. Further knowledge concerning the properties of the water and the workings of the tap may guide him to place a glass under the tap, set the latter running, and so on until such time as he has drunk the water, and thereby satisfied his thirst. In this simple example, it can be seen that the objects of the man’s knowledge are situations that exist in the world surrounding him, situations such as ‘there is water in the kitchen’. It is through being directed by knowledge of these situations that the man brings about non-accidental environmental effects with his bodily movements, effects, that is to say, that reflect the combining of his bodily drive state with his knowledge relation to the objects about him that can, potentially, be used to reverse that state.

In this we see the connection between cognition and the non-accidental environmental effects of bodily movements, as distinct from the accidental environmental effects. The latter, to be sure, are no less caused by the man’s bodily movements than are their non-accidental counterparts; what distinguishes them is that at no time did consideration of them enter into the thought-drive state combination that propelled the man’s movements in the first place. Clearly, then, there is a definitional connection between the non-accidental environmental effects of bodily movements and cognition that is lacking between accidental environmental effects and cognition, and, as will be recalled, the difference between the causation of those effects that are non-accidental and those that are accidental is what distinguishes an organism’s behavioural processes from other causal processes which, while involving the organism in some way or other, do not form part of its behavioural repertoire.

So where, then, is behaviour located? Dretske (1988) was certainly correct in claiming that the endpoint of behaviour resides in the environmental consequences of bodily movements, but his identification of the beginning point of the behavioural sequence as residing *within* the organism was mistaken, as seen by the fact that the cognitions which are required to distinguish non-accidental from accidental environmental effects, and thereby behavioural from non-behavioural sequences, are relational situations involving both the organism that knows, and the environmental situations that are known, that is, situations that extend beyond the organism. This being the case, behaviour both begins
and ends outside of the organism, even though it must at some intermediary stage involve processes that are internal to the organism. Accordingly, behaviour is located in the environmental system which contains the organism, and with which the organism interacts.

5.2.3 The role of cognitive relational situations in behaviour-causation

In accordance with the arguments of the previous section concerning the character and location of behaviour, it is clear that a theory of mental causation must show not how thought causes bodily movement, but how thought causes the thinking organism to produce non-accidental effects in the environmental system in which it is situated. Furthermore, critical to this enterprise is the demonstration that the connection between thinking and the production of such effects is an efficient-causal one.

It was argued in Section 4.3 that causation always takes place within a causal field. Causal fields, it will be recalled, are sets of conditions upon which causal events impinge, and within which effects are produced. What, then, constitute the causal fields upon which mental events impinge, and within which non-accidental environmental effects are produced?

For the most part, this question has been answered already. Behaviour, we have seen, is located in the environmental system containing the organism. The causal field within which behaviour occurs, therefore, is constituted by the organismic and environmental variables with which mental events such as ‘coming to know that p’ interact. Of the former, perhaps the most important for behaviour-production are the primary drives (see Section 4.4.1). As Maze (1983) has demonstrated, the activation of an infant’s primary drives is characterised by the performance of simple innate actions that, environmental conditions permitting, are consummatory, that is, sufficient for the reversal of the conditions stimulating their own activity. The development of these simple innate actions into more complex behavioural actions comes about with the maturation of the organism in terms of both its motor control and, importantly, its recognition of the universal characteristics of the environmental situations surrounding it. Such universal characteristics are the specific environmental variables that comprise the component of the causal field of behaviour that is not constituted by the organism itself.
To illustrate this latter claim, consider once more the example of the previous section whereby a thirsty man drank a glass of water and in so doing reversed his thirst-condition. How could knowing that the fluid in the glass was water have caused the thirsty man to have drunk it? The answer to this question is that, in and of itself, such knowledge would not have sufficed to have directed the man to drink. What is further required for knowledge of this situation to be efficacious is that, in identifying the fluid as water, the man also recognises it as belonging to the type of thing that cures thirst. Of course, there is no necessity by which knowing the first fact, that the fluid was water, would have led to knowledge of the second fact, that it was the type of thing that cured thirst. All that is being claimed is that if knowledge that the fluid was water were to be efficacious, then knowledge of water’s membership in the class of things that cure thirst would also have had to have been known.

Notice here that the specific universal characteristic of the water that requires recognition is intimately connected with the type of environmental effect that would be non-accidental if it were produced. Of course, the water has universal characteristics that are unrelated to its membership in the class of things that cure thirst, the recognition of which would be irrelevant to whether or not the man was caused to drink the water. What is essential, therefore, for the production of non-accidental environmental effects, is that the behaving organism recognises the connection that exists between the situations that form the objects of its thoughts and the type of non-accidental environmental effects that are produced. In other words, that it recognises the membership of the situations about which it thinks within broader classes that are related to the kinds of effects that reverse the factors that are stimulating its drive activities.

The causal field of behaviour, therefore, is comprised of the organismic variables, most notably the primary drives, that motivate the organism to move, as well as the known universal characteristics of the situations about which the organism thinks. This is the field into which mental events of the sort ‘coming to know that $p$’ are introduced, and, when they are favourable, these causal field conditions combine with mental events in such a way that non-accidental environmental effects are brought about. Of course, the causal field conditions might be such that the introduction of a mental event amidst them results in something very different from the effect (thirst reversal) that featured in
the example above. If, as in the above example, the man came to know that the fluid in the glass was water but some of the organismic field conditions differed, such as would occur if he were paralysed or non-thirsty, then it could not be expected that this same type of causal event would produce the same kind of effect (drinking and the quenching of thirst), because production of the same kind of effect is only guaranteed within the same kind of causal field.

Equally, if some of the environmental variables differed from those of the original causal field, if, for example, the universal character of the situation believed in (that ‘this fluid is water’) had been mistaken, as indeed is the case in the present example which should really read ‘pure water cures thirst’, then drinking could occur just as if an actual universal fact had been recognised, but the ensuing environmental effect may well in this instance be an accidental one (consider the case where the thirsty man drinks of salt water, only to discover his condition is made the worse for his action), and the man’s ‘behaviour’ could no longer be classed as ‘quenching thirst’.

Interestingly, misadventures such as these, when they do not kill us, force upon us the recognition that either the beliefs that are introduced into the causal field (in the present example, that ‘this fluid is water’) or the supposedly universal characteristic relating to the introduced belief (here, that ‘water cures thirst’) is incorrect, and, if it be the ‘universal’ that is thought to be the problem, future occurrences of beliefs such as that ‘this fluid is water’ will no longer suffice to bring about the type of behaviour that they once did, because those beliefs will then interact with a causal field different from the original one, one from which the necessary belief in the character of the universality of the situation observed will have been removed or refined.

Through appreciating the role that recognition of situations’ universal characteristics has for an organism as it learns about its environment, we better understand why seemingly novel situations, espying a hitherto unnoticed glass of water for example, can instigate an organism to interact with that situation in a way that serves its needs, that is, deactivates its primary drives. The occurrence of a behavioural sequence, such as one in which a thirsty organism perceives a glass of water, drinks the water, and is relieved of its thirst, is therefore a sequence in which perception of its adjacent environment together with knowledge concerning situations of the type perceived and their lawful
connections to certain environmental effects interact with the organism’s current drive
state, producing its bodily movements. The success of these movements in deactivating
the drives confirms both the newly formed belief and the already existing universal one,
whilst failure to reverse the drives provides a source of doubt concerning one, the other,
or both of them.

Furthermore, the experience of success in reversing an activated drive state through
acting in certain ways upon an environmental situation that is perceived reinforces the
behaving organism such that if it finds itself in the same type of drive state and the same
type of environmental situation again, a relevantly similar behavioural sequence will be
instigated. Thus, as the organism matures and comes to recognise the great variety of
classes or types to which any particular object belongs, it is enabled to act upon that
object in accordance with the capacities of members of that class of objects to be
manipulated in ways that may be of benefit to itself.

In summary, the causal field within which behaviour is produced by mental events is
comprised of both organismic and environmental variables. When these causal field
conditions differ, when, for example, drives are deactivated or universal facts remain
unknown, then the introduction of new beliefs will differ in their effects from those
produced when drives are activated, or universal facts are known. In other words, it is
an organism’s perception of environmental objects that, causal field conditions being
favourable, causes it to produce non-accidental environmental effects, or, more
concisely, to behave.

A word remains to be said confirming the commitment of the theory outlined above to
determinism, because examples whereby a person’s thoughts and state of drive
activation combine with each other and thereby produce bodily movements that are
effective in reversing the original drive state are often treated in the mental causation
literature as evidence for the purposivity of human behaviour. Teleological accounts of
why the man featured above moved towards the kitchen, sought the glass, filled it and
drank seem to many authors to be the appropriate way of explaining this sequence; the
man did what he did in order to cure his thirst. But, of central importance to the Direct
Realist’s characterisation of behaviour as a process involving environmental effects
that, because they were caused by beliefs concerning them, are non-accidental, is the
fact that this causal process, like any other, is non-purposive (the centrality of
determinism to Direct Realist theories of mental causation was introduced above, see
Section 4.3.4).

In other words, whilst agreeing with purposivists that people behave because of the
promptings of their drives, and the contents of their beliefs, at no point is it suggested
by Direct Realists that organisms choose to act in accordance with their beliefs in order
to satisfy their drives, although they may just as well have done otherwise. Rather, it is
hypothesised that an organism’s recognition of the relevant universal character of its
beliefs about its environment is a sufficient condition, within the right kind of causal
field, for the causation of a behavioural process. Rather, then, than acting upon its
beliefs in order to satisfy its drives only when it chooses to do so, an organism is
compelled by its perceptions of the environmental situations surrounding it, as well as
by its knowledge relations concerning the universality of these situations, to act when
stimulated by the activity of its primary drives.

It should be noted at this point that much of what is being suggested in the present
section bears a great deal of resemblance to what Maze (1983) has argued for, and,
indeed, has been influenced greatly by what he has discovered, particularly as concerns
the primary drives. Nevertheless, as was concluded in Section 5.1, Maze’s account does
not demonstrate the role played by cognitive relational situations in the causation of
behaviour, opting instead to focus upon the role of intrinsic organismic states, more
specifically brain states, in such a way that the externality of thought is ignored, and the
relatum term of the cognitive relation, the environmental situation, seems not to be of
causal relevance to what the organism does. Rather, then, than focus solely upon the
referent term of cognitive relational situations (i.e., brain states) when giving an account
of mental causation, the present account emphasises the fact that it is the organism’s
entering into cognitive relational situations with external environmental situations that
is of relevance for initiating behavioural sequences. This, unlike the brain state
depiction of mental causation, accounts for why the object of thought, more particularly
the structure and content of the environmental situation that is known, is of central
relevance to the direction that behaviour takes, in a way that the mere presence of a
certain internal brain state or process could not. And this brings us around once more to
the main problem that Section 5.2 has been working towards resolving, the question of
whether cognitive relational situations can be causally efficacious in behaviour-production.

In affirmation of the causal role played by cognitive relational situations in behaviour-production, it has been suggested that when such relational situations are introduced into the appropriate causal fields, which are comprised of organisms and certain variables of the environmental systems in which they are contained, behavioural sequences are initiated. Evidence supporting this suggestion has included appeal to the general character of causes as events in which spatio-temporally extended situations impinge upon causal fields, and the fact that both impinging event and causal field involve properties and relations, both of which appear to contribute to the event’s causal efficacy within those fields. More specific evidence that it is not only possible for cognitive relational situations to be causally implicated in behavioural processes, but that such processes could not occur unless organismic knowledge of environmental situations obtained, was found upon analysis of the centrality of non-accidental environmental effects to the occurrence of behaviour, and the definitional role of cognition’s causal connection to these. But does evidence of this sort satisfactorily answer the question that has been posed?

One reason why it might be supposed that something is wanting in the above account of mental causation is that it appears to be very broad; perhaps lacking in the detail that an account of neural firings, synaptic processes and so forth could provide in relating what occurs when organismic perception causes bodily movements and non-accidental environmental effects within an appropriate environmental system. In response to concerns of this sort, it can only be replied that because cognitive relational situations cannot be divided into smaller components without the loss of their characters, no more minute account of the mechanism by which perceiving one’s environment triggers behavioural processes is available. Identification of the brain states and processes that are necessary for cognition to occur may well lead to discoveries concerning connections between these and the motor responses that are caused by them, but discoveries of this sort are not the least bit illuminating when it comes to the question that was originally posed, that of how thinking causes behaviour, because they are not concerned with thinking and behaviour themselves. Naturally, then, rejection of the above account of mental causation on the grounds that it does not address neural
mechanisms, or anything else with which it was not originally concerned, would be unfair and counter-productive, the latter because it would evince a confusion between cognitive relational situations in their entirety, and just one of their terms (the brain processes).

Furthermore, if fault is to be found in the identification of the kinds of events that cause behavioural processes in so far as these are indivisible, then fault is to be found in any causal sequence that one cares to name. What more could there be to explaining causal processes than identifying the kinds of initiating events that reliably bring about certain kinds of effects within certain kinds of causal fields? In identifying these events, we specify the efficacious properties and relations contained in the situations involved in them, and if it is to be argued that this does not suffice for causal explanation, then one is entitled to ask what does. The idea that a more minute account is always possible, that one can look to ‘lower’ levels for a better explanation, is founded upon the mistaken notion that there is no difference in identity between a unitary complex situation and some part or parts of that situation, and therefore that elucidation of the latter will provide elucidation of the former. However, attempts to ‘break down’ complex situations into their elements, for example, to examine how societies work in terms of individuals or how individuals work in terms of neurons, inevitably alters the character of the investigation with which one began, and thereby undermines the investigation itself.

To demonstrate this, consider one of the common kinds of thought experiment that is popular in the mental causation literature, a situation whereby my brain is somehow (not through perception) made to function in a way that is indistinguishable from the way that it functions when I am involved in perceiving a glass of water. If, the argument runs, functioning in this manner suffices for my reaching out and drinking the glass of water in this instance, then surely in other instances, those in which perception of the glass of water occurs, the efficacious events are likewise simply those involving my neural activity, and not the entire perceptual relational situations that extend beyond me.

Compelling as arguments such as these may at first appear to be, they do not bear close examination. While it cannot be denied that there may be any number of ways of making my arm extend, grasp a glass, raise it to my lips and so forth, it cannot be
maintained that in each of these scenarios I am involved in bringing about non-
accidental environmental effects, that is, behaving.

Consider once more the example used to demonstrate the difference between a rat
pressing a lever, and an experimenter placing the rat’s paw on the lever, and thereby
caus ing the lever to be depressed (see Section 5.2.2). While in both instances the paw
moves and the lever goes down, only the first could reasonably be identified as an
instance of rat behaviour. Similarly, when my brain is made to function, by some
artificial means, in a way that is similar in all relevant respects to the way that it does
when I perceive a glass of water, it may well be the case that my body is made to move
in a way that results in the water being poured down my throat, but this effect could no
longer be regarded as non-accidental any more than could any other effect that
occurred, because non-accidental effects of bodily movements are distinguished from
accidental effects by the organism’s foreknowledge that bodily movements of the type
made bring about effects of the type effected, and no such foreknowledge plays a role in
the artificial brain stimulation scenario.

Consequently, the seemingly similar scenarios, the first in which perceiving a glass of
water makes me drink, and the second in which artificial brain stimulation causes my
body to move in such a way as to grasp the glass and pour the water down my throat,
are no more of a species than are the situations wherein one student excels in an exam,
and another student is awarded the same mark for copying the answers of the first.
Accordingly, because different kinds of causal sequences are involved, the explanations
that are appropriate for them will of necessity differ, and it could not reasonably be
expected that the explanation that suffices for one will be of any use in understanding
the other. The objection that the identification of the mental events that cause bodily
movements and non-accidental environmental effects is of too broad a character,
therefore, is one which should itself be objected to. The answer, then, to the question of
whether cognitive relational situations can be and are causally efficacious, is an
unqualified ‘yes’.
Section 6: Summary and Conclusions

Many of the controversial issues with which psychology is concerned, and which give rise to an abundant variety of theories of mental causation, concern the question of the place of human mentality in the natural world. Donald Davidson’s theory of Anomalous Monism suggests that the character of human mentality is such that it, alone, exhibits rationality, and this not contingently but necessarily. Related to the proposed rationality of human mentality is the suggestion that mental acts proceed independently of governing laws, allowing them freedom to vary with an independence that is absent from the lawfully constrained workings of physical processes.

Davidson’s commitment to the rational character and anomalous proceedings of mental events is founded upon an approach to event ontology whereby events are deemed to exist and function independently of the properties by which they may be described. Attributions of properties to events are nevertheless seen to be necessary for the classification of events into types, such as mental types, and the processes by which attributions are made are understood to be constrained by the *a priori* dictates of theories, such as the Unified Theory, that prescribe the characters of the phenomena with which they are concerned.

The distinction that Davidson proposes to exist between event ontology and conceptualisation, that is, between how events exist in and of themselves and how they are conceptualised and described, is reflected in his account of the distinction between the processes of causation and the character of causal explanations. Causal processes, Davidson maintains, involve events themselves independently of any properties that may or may not be attributed to them. Causal explanations, on the other hand, are intimately concerned with, and appropriately constrained by, the characters of the properties that the causal events are attributed with. Accordingly, explanations of causal occurrences in which mental events are involved must conform to the pre-existing assumptions of the theory by which the events were originally judged to be mental. These assumptions allow for mental causal explanations to differ from physical causal explanations, not only in that the former are deemed to be exempt from strict laws, but also in that they are understood to be compatible with the conceptions of free will and goal-directed or teleological causation that are embraced by practitioners of folk
psychology, particularly where such conceptions are those best suited to conjuring a picture of rational human thought and action.

The key, Davidson argues, to recognising the compatibility of the lawful character of causation and the anomalous character of mental events lies in one’s prior recognition that it is events, not the properties with which events are attributed, that participate in causal relations. Once this is accepted, it becomes evident that mental events are physical events, because all events that are describable within the mental vocabulary are likewise describable in the physical vocabulary, and, therefore, that while explanations of causal processes that are couched in mental terms will necessarily be non-strict and possibly even non-efficient, explanations of those same causal processes in physical terms will be suitable for refinement into exceptionless causal laws.

In proposing the identity of mental events with physical events, or, more accurately, the amenability of particular events to description in both the mental and physical vocabularies, Davidson takes care to avoid suggesting that the two modes of description are themselves lawfully connected, realising that such an admission would inevitably undermine his assertion that mental events are anomalous. Instead, he argues that the connection between mental property attributions and physical property attributions is a species of weak supervenience, whereby the physical attributes of others are interpreted in accordance with their overall sets of mental characteristics in such a manner as to produce a judgement and associated mental property attribution that accords with the complete picture of the others’ psychological make-ups in the most internally consistent way possible. The mental property that is attributed to the person being interpreted is, Davidson maintains, of a tentative character, and, furthermore, it is perfectly feasible that alternative property attributions that are incompatible with each other may each prove to be equally consistent with a rational overall interpretation of the person’s thoughts, deeds and words, leaving the interpreter with no in-principle way of deciding between them.

Davidson’s theory of Anomalous Monism has been the subject of criticism within the mental causation literature on two main grounds. Firstly, it is argued that Anomalous Monism is compatible with, and perhaps entails, mental epiphenomenalism. But if, as Davidson suggests, it is not properties but events that are causally efficacious, then
arguments suggesting that mental events are epiphenomenal cannot be sustained, whereas arguments suggesting that mental properties are epiphenomenal are of little relevance.

Secondly, it is argued that Anomalous Monism undermines rationalisations, that is, causal explanations couched in mental terms, because of its refusal to recognise the existence of strict psychological laws. This second line of criticism is more problematic for Davidson than is the first, because, as he seems to recognise, it is only if there exist lawful connections between properties that causal explanations can be formulated, and it is not feasible, as he had maintained at first, that lawful connections between physical properties could be utilised in psychological explanations because what is required are lawful connections between the types of properties featuring in the explanations, and these are mental.

While criticisms such as these are important for identifying some of the limitations that subscribing to Anomalous Monism brings, they are rendered somewhat otiose by the consideration that only if Davidson’s theory were logically sound, which it is not, would they then be necessary to show that it fails nonetheless as a theory of mental causation. That Anomalous Monism is not sound is made evident by examination of its underlying assumption that there exists an ontological-conceptual distinction between events and properties according to which events alone exist independently of human concerns while properties exist only insofar as they are conceptualised and described in the service of human theorising and explanatory practices. As consideration of the logic of symbolism shows, nothing exists by virtue of being described, rather, successful description requires that the thing being described have an existence that is independent of how, and if, it is labelled. Additionally, Davidson’s conception of events as somehow having an existence that is independent of properties, an existence that is without universal features, is demonstrably false. Like objects, events have both particularity in a locational sense, and universality in that they are comprised of properties or features that characterise other objects and events as well. Furthermore, these properties, like the objects and events that they characterise, are mind-independent, requiring neither conceptualisation nor description by human beings for their on-going existence.
Davidson’s proposal that mental characteristics are the products of human theorising and descriptive practices evinces an anti-realist stance towards human mentality, one that divorces thought and its associates (such as meaning) from the natural world. The suggested dependence of one’s mental phenomena upon the interpretative acts of others is founded upon an elaborate although unsustainable system of proposals according to which the character and content of what one believes and desires, says and does, are denied the permanence that characterises other worldly states and events once they have come to pass. The notion that mental acts exist only if they are judged to do so, and only in the manner in which they are judged, robs them of anything more than a conceptual fictional status that is incompatible with their having causal efficacy. Therefore, far from proving to be a viable theory of mental causation, Anomalous Monism serves only to demonstrate the multitude of problems that arises when one proposes that, rather than be of the natural world, mental phenomena have a mind-dependent conceptual existence.

By recognising that cognition is a spatio-temporally extended relation that obtains between organisms and environmental situations, the theory of Direct Realism takes as its starting point the view that human mentality forms part of the natural world. Accordingly, the extension of Direct Realism into a theory of mental causation requires acceptance of the fact that the characteristics of causation in the natural world apply to causal processes in which mental events are involved, and therefore that the latter are in no sense exempt from the conditions of necessity and sufficiency that feature universally in efficient causation.

The conflict which some hermeneutically oriented theorists imagine to exist between rationality as an inherent feature of human thought and action and the deterministic principles of causation, a conflict which suggests to them that mental causation might proceed according to teleological or other non-efficient principles, does not arise for the Direct Realist, who understands human rationality to be no more than evidence of organismic sensitivity to the logical way in which the world is structured. Accordingly, rather than conclude that human action is inherently rational and therefore necessarily purposeful because it serves the needs of those engaging in it, a Direct Realist account of mental causation aims at demonstrating that behaviour-production is an efficient causal process whereby the organism interacts with the world around it when and only
when sufficient causal requirements are met, these including organismic, environmental and cognitive factors.

In his account of the functioning of the primary drives, John Maze identifies the organismic conditions that stimulate immature organisms to perform simple innate actions, and more mature organisms to perform more complex behavioural acts. The latter actions, however, require also that the organism perceive facts about the environment in which it is currently located, as well as the connections between the universal features of these facts, and the kinds of environmental effects that they may bring about if interacted with in appropriate ways (most notably through bodily movements). In this latter requirement we discern not only that cognition is causally necessary for behaviour to occur, but also that cognition must be a form of connection between organisms and the structure and contents of their surrounding environments, and therefore external to the organism itself, just as Direct Realism suggests.

Accordingly, the externality of cognitive relational situations that is proposed by Direct Realism, rather than proving mysterious, is in keeping with the fact that cognition is not an internal organismic occurrence, but a form of interaction between the organism and the world at large. Furthermore, as the consideration that all causal events are spatio-temporally extended situations shows, the incapacity that many authors assume to prevent cognitive relational situations from acting as causes because they are relational and therefore extended in character is illusory. Coupled, then, with the evidence that there is no obstacle to relational situations acting as causes, the Direct Realist view of cognition provides a secure foundation upon which to establish a logically sound theory of efficient mental causation.

In conclusion, one of the most important contributions that Direct Realism has to offer to psychology is the denial of the existence of an ontological boundary separating human mentality and the natural world into mental and physical ‘realms’, together with the positive assertion that there exist connections between all parts of nature, including humans, some of which have a cognitive character. By recognising and demonstrating that human mentality forms part of the natural world, and that cognition is relational in character, Direct Realism provides a basis for the resolution of many of the issues, including that of mental causation, with which psychology is currently concerned.
References


Watson, J.B. (1913) Psychology as the behaviorist views it. Psychological Review, 20, 158-177.
