

DO CAUSES NEED TO MAKE THEIR EFFECTS PROBABLE IN ORDER TO EXPLAIN THEM? THE TENSION BETWEEN N1 AND E5 IN CRAVER'S MECHANISTIC MODEL OF EXPLANATION

Ben Crisp

Carl Craver (2007) proposes a mechanistic model of explanation in science motivated by a desire to intervene, as exemplified by explanations in neuroscience which in his opinion are motivated by the desire to bring the central nervous system under control. In his discussion of causal relevancy conditions of mechanistic components Craver asserts that a cause need not make its effect probable in order to explain it (condition E5). Although this is supported by some interpretations, Craver's own is highlighted by his appeal to an example in neuroscience of the apparently stochastic nature of neurotransmitter release events. I propose the view that this interpretation of such a causal relevancy condition is contradictory to Craver's own proposed first norm of explanation (norm N1): that a good explanation must fully account for the explanandum phenomenon including varied manifestations of the phenomenon. By defending Craver's explanatory motivation of intervention I demonstrate that the attributed cause of a phenomenon must necessarily make the phenomenon probable in order to effectively explain it.

Keywords: mechanistic explanation, scientific explanation, causal relevancy, Craver

INTRODUCTION

Carl Craver's (2007) mechanistic model of explanation offers a convincing alternative to Carl Hempel's (1966) classical covering law model of explanations in science. Rather than modelling explanations as arguments, Craver outlines the norms of explanation evident from examples in neuroscience to demonstrate that good explanation (in his opinion) is the description of the organisation of entities and their activities comprising mechanisms that exhibit the explanandum phenomenon. Drawing on the statistical relevance model of Wesley Salmon (1989), Craver constructs a workable normative account for explanations in neuroscience which are, as he believes, motivated by the desire to bring the central nervous system under control.

Citing Salmon, Craver (2007) asserts that in determining the causal relevance of the activities of components a cause need not make its effect probable in order to explain it (E5 of Craver's constraints of causal relevance). Craver illustrates this point by reference to the example of action potentials and calcium influx explaining neurotransmitter release, despite only 10 -20 percent of the precipitating conditions actually resulting in release events. However, the first of Craver's norms of mechanistic explanation (N1) is the necessity that an explanation fully accounts for the explanandum phenomenon, including precipitating conditions, manifestations, inhibitory conditions, modulating conditions, non-standard conditions and by-products.

I intend to demonstrate that if explanations are motivated by a desire for control then a normatively good explanation must necessarily incorporate enough information to demonstrate how, given a discrete outcome (such as a neurotransmitter release event), its causes (the discrete values taken on by variables described within the mechanistic explanation) made it the probable ($P > 0.5$) phenomenon over a probability distribution of possible phenomena. If this is impossible, due to the stochastic nature of constituents within the explanation, then the control-motivated explanation should not be considered good.

WHY EXPLAIN?

The epistemology of explanation is a topic of much debate in the philosophy of science. An adequate account of the norms of explanation remains elusive; however, in *Explaining The Brain* (2007) Carl Craver has recently published an attempt to describe his own account influenced by explanations from neuroscience. Craver claims that good explanations of phenomena in the central nervous system (CNS) within neuroscience are descriptions of mechanisms, comprising the organisation and activities of entities such that, collectively, they exhibit the *explanandum phenomenon*. Furthermore, explanations in neuroscience are motivated by utility. It is their usage – whether for purposes of intervention upon processes in the CNS, the development of treatments for failures or abnormalities of the CNS or the development of external computational systems modelled on the CNS – that shapes the nature of the explanations sought by observation and experimentation in neuroscience. Craver sums this up nicely: “Explanations in neuroscience are motivated fundamentally by the desire to bring the CNS under our control” (Craver, 2007, p. 160).

One example proposed by Craver of a mechanism in neuroscience is the action potential – an electrical impulse that travels down the axon (a membranous tube filled with ion-rich aqueous solution) of a neuron to the synaptic terminal, where it affects the chances of an alteration in the post-synaptic cell's electrochemical activity (Bogen, 2005, p. 402). The description of the mechanism underlying this phenomenon is a complicated compilation of models, formulae and other forms of illustration based on the work of, amongst others, Hodgkin & Huxley (1952), C.M. Armstrong (1981) and Bertil Hille (1992). It describes the progressive transmembrane movement of various ions such as Na^+ in the solution within and without the axon via ion-specific α -helix "ion channels" in the membrane which are gated by particular ball-and-chain-style protein structures (Craver, 2007, pp. 114-122). Another example used is the phenomenon of neurotransmitter release at the synaptic cleft, preceded by the action potential (but not predicted by it), explained in part by the influx of Ca^{2+} ions.

These examples from neuroscience are not sufficiently-well explained, Craver argues, until descriptions of their mechanisms (the organisation of entities and activities exhibiting the phenomena) satisfy at least the following norms:

- N1) The explanandum phenomenon must be *fully* accounted for (including precipitating conditions, manifestations, inhibitory conditions, modulating conditions, non-standard conditions and by-products);
- N2) Entities of the mechanism must be robustly detectable, able to be used for purposes of intervention and physiologically plausible;
- N3) Activities of entities must satisfy conditions of causal relevance that rule out mere time-courses, effect-to-cause pairings, effects of common causes and irrelevant causes;
- N4) The organisation of entities must involve relevant active, spatial and temporal interactions distinct from mere aggregate, morphological or taxonomical explanations; and
- N5) Constitutive relevance must be demonstrated by an appeal to mutual manipulability (e.g. empirical evidence that an ideal intervention on a component's activity in the mechanism alters the exhibition of the explanandum phenomenon *only via* the change in the component's activity).

(Craver, 2007, p. 161)

Each of these norms corresponds to the motivation of explanations in neuroscience, and each is necessary to avoid the pitfalls of classical approaches to explanation such as the Covering Law (CL) model of Carl Hempel. The CL model proposed explanations as deductive arguments, where the explanandum phenomenon is deduced as the conclusion of a logical series of true premises at least one of which is a law of nature (Hempel, 1966). Craver builds upon the work of Wesley Salmon to demonstrate that, rather than reducing phenomena to natural "laws" and logic, good explanations describe how phenomena are causally positioned within the observed world (Craver, 2007, p. 21).

A common argument against the CL model is that by relying wholly on the validity of reasoned argument extraneous and irrelevant information may be included in an explanation without weakening its strength. For Salmon, this marks a clear weakness since the strength of an explanation is directly related to the relevance of the included information (Salmon, 1989, p. 102). It is uncontroversial that the blessing of neurons with holy water will not be explanatorily relevant to their producing action potentials despite the possible empirical evidence showing that stimulation of blessed neurons produces action potentials (from an example in (Craver, 2007, pp. 36-37)). Similarly, there are many correlations between possible variable-values of constituents (the range of realisations of activities and entities) of a mechanism and the nature of the explanandum phenomenon that are not explanatory in themselves. Craver cites Salmon's own constraints upon assertions of causal relevance within explanation:

- E1) Mere temporal sequences are not explanatory (temporal sequences);
- E2) Causes explain effects and not vice versa (asymmetry);
- E3) Causally independent effects of common causes do not explain one another (common cause);
- E4) Causally irrelevant phenomena are not explanatory (relevance); and
- E5) Causes need not make effects probable to explain them (improbable effects).

(Craver, 2007, p. 26)

The first four of these constraints are demonstrated by Craver fairly convincingly in various examples both from neuroscience and thought experiments. The rooster's infallible crowing

at dawn does not explain the sunrise (E1). The length of a flagpole's shadow does not explain the flagpole's height (E2). Falling barometer levels do not explain thunderstorms (E3). Blessing neurons does not explain their action potentials (E4). It is the final constraint of improbable effects (E5) that needs further discussion.

FULLY ACCOUNTING FOR THE EXPLANANDUM PHENOMENON (N1)

Not all action potentials and Ca^{2+} influx result in the release of neurotransmitters. What does a description of this mechanism actually describe then? Salmon's approach to causal relevance is an appeal to *statistical relevance*. Whilst it does refer to probabilities, it necessitates only that an attribute X is statistically relevant to an attribute Y if and only if the probability of Y conditional to its membership to population Z *and* presence of X is *different* to the probability of Y conditional to its membership to Z alone (Salmon, 1971). Therefore the attributes of action potentials and Ca^{2+} influx are statistically relevant by virtue of their alteration of the probability of release of neurotransmitters. In fact, neurotransmitter release is *unlikely* to occur following an action potential inasmuch as empirical testing has demonstrated only 10-20 percent likelihood (Craver, 2007, p. 26).

However, it would appear that in such a case the explanandum phenomenon has not been fully accounted for in accordance with (N1) of Craver's norms. Craver himself stresses the importance of knowledge of "inhibiting conditions – that is, the conditions under which the phenomenon fails to occur" (Craver, 2007, p. 126). Furthermore, the release of neurotransmitter is sometimes known to occur independently of action potentials; prompting Craver to assert that "in some cases there are two independent possible explanations" (Craver, 2007, p. 24). Yet for an explanation to fully account for the explanandum phenomenon it seems unclear how there may be "two possible independent explanations". Instead, it seems to make more sense to describe a mechanism that may have mutually independent components which nevertheless are organised and in activity to exhibit the explanandum phenomenon.

When discussing improbable effects Craver refers to an example classically offered by Michael Scriven of the case when untreated syphilis results in general paresis. Despite only 20 percent of cases of untreated syphilis infections manifesting in general paresis, syphilis is cited as the "cause" of paresis. Craver offers no description of the mechanism of this

example, instead using it as a case where a cause fails to make the effect probable. A similar example, typically used in favour of probabilistic causation, is the proposition that smoking is a "cause" of lung cancer. Whilst smoking is not invariably followed by the development of lung cancer empirical studies have demonstrated the regularity with which the former increases the probability of the latter – thus suggesting a cause-effect pairing based on probability (U.S. Department of Health and Human Services, 1982). This pairing does not explain *how* smoking causes lung cancer; it merely offers good reason to more deeply investigate the suggestion that it may.

Phenomena such as those in the examples above are typically classified as "stochastic" because of their indeterministic nature. The inability to determine the outcome of an event is a major focus of microphysics since particles at a subatomic level often have no definite location or trajectory. Thus the movements of such particles, such as in the decay of the nucleus of radioactive isotopes, are often unable to be predicted with much certainty. At a macrophysical level, however, events such as complex weather patterns or socio-economic fluctuations are labelled stochastic when the variables required to determine the outcome of a causal sequence are either unknown or incalculable. These variables are either *outside* the scope of the suggested explanation of the phenomena, or they are unaccounted-for variables relevant to the sub-mechanisms of processes *within* the observed phenomenon's mechanism. Despite the fact that observations of macrophysical events can eventually be reduced to observations of subatomic particles that may indeed behave stochastically, much regularity is evident at a macrophysical level of observation. Most scientific experimentation is aimed at identifying and classifying these regularities. Scientifically "proven" regularities or theorems, beyond the mathematical consistencies of some physics, are only those that have successfully been realised in a controlled observation over a sufficient number of trials and are yet to be "disproved" by a recorded case of the theorem failing. Whilst fundamentalist physicists and indeterminists argue that it is impossible to predict outcomes of events with 100 percent certainty, scientists base most theories on thresholds of acceptable probabilities.

This rough description of how realised cases inform the explanations of general cases indicates that there is a distinction between explaining a mechanism extant in the structure of the world and explaining instances of that mechanism. It may be useful to examine a case that exhibits apparently stochastic behaviour.

A MECHANISM TO EXPLAIN DIE-THROWING

Suppose we tabulated the variables relevant to the constituent entities, activities and organisation of a mechanism of die-throwing:

THEORETICALLY MEASURABLE MECHANISM ENTITIY VARIABLES

Angle of impact relative to three planes (AI1, AI2, AI3)

Starting position relative to table surface plane in three dimensions (X, Y, Z)

Trajectory of throw relative to table in three dimensions (T1, T2, T3)

Revolutions made before impact on three axes (R1, R2, R3)

Velocity of spin on three axes (A1, A2, A3)

Time of travel (TT)

Velocity of impact relative to three planes (VI1, VI2, VI3)

Bounces (*b*)

Bounce trajectory over *b* changes in acceleration (BT1.1, ... ,BT*b*.1)

Die elasticity (E)

Table energy absorption (TEA)

.
. .
. .

Table topography (S1, ... , Sx)

Whether or not these values are measurable by the human eye is irrelevant. At least theoretically, there are numerous variables in this example that could be quantified with a range of scientific equipment such as high-speed cameras, microscopes, x-rays, etc. Of course, variables could continue to be measured and added to this list indefinitely. The depth of detail of a qualified "good explanation" (but not necessarily a "complete" explanation) must necessarily be limited by the context within which it is being examined. The point is, a *sufficiently* detailed account of a die-rolling mechanism comprising enough variables will presumably produce a function upon those variables resulting in a probability distribution of roughly one in six, given the range of *all* possible values for each variable. Consequently, given a set of *discrete* values for the mechanism variables \mathcal{S} [$A1 = 0.3\text{m/s}$, $R1 = 12.3$, ... , etc] one discrete outcome will be determined by the function.

A correlating function itself is not sufficient to explain a phenomenon. The Hodgkin and Huxley equation describing the various characteristics of the action potential is not held to be explanatory; nor was it believed to be so even by Hodgkin and Huxley themselves (Hodgkin & Huxley, 1952). The ability to predict the characteristics or occurrence of a phenomenon is not sufficient or equivalent to understanding it. Craver demonstrates this fact by pointing out that the height of a flagpole can be determined as a function of the length of its shadow and the elevation angle of the sun, despite the fact that neither of these variables is causally relevant to the "phenomenon" of the flagpole having a certain height (and thereby not satisfying (N3)) (Craver, 2007, p. 25). However, the Hodgkin and Huxley equation (ignoring its inaccuracies and the fact that it simplifies and aggregates a number of variables for a moment) is a part of a good explanation of the mechanism underlying the phenomenon of the action potential, since the values it is a function of represent the organisation of and the characteristics of the entities and activities of the mechanism.

What part of the good explanation is an equation such as the Hodgkin and Huxley model, or indeed a function on the variables of die-rolling? In essence, the equation is an explanandum phenomenon unto itself – it is an observable organisation of values (representative of characteristics of physical things) extant in the causal structure of the world. Empirical testing is needed to link "filler terms" (such as unknown constants and variables like those in the Hodgkin and Huxley equations) with robust entities to satisfy (N3). A function on the variables of die-rolling may describe a correlation of macrophysical values that are "reducible" to relative activities at a microscopic, molecular, atomic or subatomic levels – just like the activities of the ions of the Hodgkin and Huxley model may be reduced to subatomic activities. They may reach irreducible entities such as gravitational and electrostatic forces. Whilst views may differ on the relative merits of reductionism, it seems reasonable to recognise correlative equations whose variables represent characteristics of mechanistic entities/activities/organisations as an important part of (N1) – that is, fully accounting for the explanandum phenomenon by describing the range of manifestations of outcomes and how they are causally related to the mechanism components.

EXPLANATION OF REALISED TOKEN EVENTS AND GENERAL CASES

Excepting a few outside cases, the possible outcomes of rolling a six-sided die will be the die landing on one of the six faces. With an evenly weighted die, over a sufficiently large trial of

throws, each outcome will be observed to occur roughly one out of six throws. This is not to say that it is impossible to throw an even die one million times and land a six every time. It is just not considered as likely as a more even distribution of results. It could be argued that this expectation of an even probability distribution is an entirely arbitrary and subjective interpretation of the outcome. Indeed it *is* subjective since the outcome is typically being recorded independently of starting conditions such as, for instance, the position, angle and velocity of the die.

What we would expect of our description of the mechanism of die-throwing is that, were we to tabulate the complete set of possible ranges for each value of each variable (to some prescribed level of detail, e.g. $S(A1) = [0.1, 0.2, \dots, 359.9, 360]$, as determined by our requirements), we would find that the discrete outcomes (i.e. landing on a 1, 2, 3, 4, 5, 6) occur with a frequency of around 0.167 (or 1/6). Empirical testing of the causal relevance of these variables would require conditions where each variable may undergo ideal interventions; a robotic arm that is capable of throwing a die with a preset weight distribution, velocity and angle of spin at a table surface of which detailed topographical information is known.

Suppose then, that a die is rolled and lands on a six. How do we explain how the die landed on the six? By reference to our empirically tested description of the mechanism of die-throwing we can explain any of the results of the throw. It seems as though the same mechanism is being described whether the die landed on a six or not. However, the explanation for why the die landed on a six should not be the same explanation as why the die landed on a one, since the events are different. There must be some further qualification that distinguishes the explanation of the realised case (e.g. the die landing a six) from that of the general case (e.g. the die landing on a face). To complicate matters further, suppose we calibrate our die-throwing robot with a value set from within \mathcal{S} that corresponds to a highly probable ($P > 0.99$) result of the die landing on a six. Then, upon activating the throw, the die lands on a one. How is *that* result explained? Is there some other "independent possible explanation"? Or is it that our die-rolling explanation is limited by the detail of variables that are considered?

THE MOTIVATION TO INTERVENE

If explanation in neuroscience is motivated by the desire to intervene in the activities and organisations within mechanisms of the CNS, as Craver suggests, then it is necessary to be able to predict the outcome of interventions. Most interventions on variables within a described mechanism will result in a change in the distribution pattern that represents observations of the explanandum phenomenon in empirical studies.

Loading a die is possible since weight distribution within a die is known to be causally relevant to the outcome of throwing it; therefore intervening to manipulate the weight distribution allows one to increase the probability of one of the discrete outcomes. However, whilst loading a die may make one discrete outcome *more* probable than the rest it may not necessarily make that outcome *probable* (i.e. $P > 0.5$). Indeed, intervening to "set" any of the *individual* variables within the mechanism description may not make any of the outcomes *probable* (since the probability distribution of the discrete outcomes may be identically even for all values of A1, for instance).

What is the difference between the explanation of why an even die landed on a six as opposed to a loaded die? The general mechanism is the same for both cases. The answer, it would seem, is that in a realised case the *actual values taken on by the variables is important*. Not everyone would agree with this. Woodward describes a similar example in which an unevenly tails-biased coin is tossed. No matter what the outcome is, if one can understand the likely case (landing on tails) then one can equally understand the unlikely outcome (landing on heads) so long as the relevant factor (tossing the coin) is understood (Woodward, 2003). Woodward uses this example to demonstrate the weakness of Carl Hempel's inductive statistical extension to his classical deductive nomological approach of the covering law model. This model proposed that explanandum phenomenon can be derived from subsumption under statistical laws of the universe. As has been demonstrated above, it is seems difficult to assert that there are object statistical laws of the universe – or if there are that we could ever possibly objectively quantify them. Instead, it would seem that all probabilities dealt with in experimental science are subjective to the range of variables we can manipulate or at least measurably observe. Woodward's example fails to realise that whilst the tossing of the coin may be the only relevant factor to achieving an outcome, is not the only relevant factor needed to explain the discrete outcomes of the explanandum

phenomenon. It is "tossing a coin" itself that needs to be described as the organisation of entities and their activities in order to effectively understand the outcomes.

THRESHOLDS OF COMPLEXITY/USEFULNESS OF MODELS AND EXPLANATIONS

Salmon's statistical relevance model suggests a condition of objective homogeneity, which holds that the cause of an effect is explained when there are no further variables which could be described in the explanation that would affect the probability of the outcome. Salmon recognises that such a condition is meaningless in many cases since the task of encompassing the full range of relevant variables impacting an outcome seems insurmountable (Salmon, 1984, pp. 53-54). Yet the scope of the explanation must be relevant to whether an explanation is good or not. Clearly we must determine some threshold of complexity or detail to an explanation if we are to sufficiently establish the norms of explanation.

Craver refers to the manifestations that explanatory descriptions of mechanisms may occur such as example explanatory texts, diagrams, models, computer simulations and so on (Craver, 2007, p. 27). The last of these examples – computer simulation – has become particularly important as a manifestation of explanation as developments in computer technology have allowed for more and more detailed modelling of complex systems. A model, by definition, is a simplified representation of an actual system – simplified usually by a limitation of variables accounted for in simulation. A model of an aeroplane in a wind tunnel will not demonstrate the full range of behaviours of the actual aircraft it represents since in reality there is an enormous range of variables that are not represented in the simulation. The closer the model aeroplane is in size to the actual aeroplane, the more accurate the results of the simulations will become – however even a full sized aircraft will not behave the same way in a wind tunnel as it will in actual flight. As more variables are represented in the model the discounted variables will become more insignificant – the energy released by the quantum decay of a particle in another galaxy will theoretically impact on particles comprising the physical form of the aircraft, but at some level of complexity we will be reasonably confident in making predictions based on a model that do not account for this incredibly infinitesimal impact.

How is this threshold of complexity to be determined? It depends on the motivation of the model, of course. If, as Craver suggests, explanations are motivated by the intention to intervene then it would be reasonable to think that a normative *good* explanation should allow (at least hypothetical) intervention with predictable results. That is, if our die-throwing robot was programmed with starting positions and instructions appropriate to a high probability ($P > 0.99$) of rolling a six, and consistently failed to roll a six in controlled experimental conditions, we should have reason to think that our description of the mechanism underlying die-rolling is seriously incomplete. Why simply put this down to irreducibly stochastic behaviour?

Furthermore, why presume that our explanation of neurotransmitter release is a good one when our explanatory precipitating conditions account for only a 10-20 percent result of release events? One answer to this question is proposed by Beck and Eccles in their quantum neurotransmitter trigger mechanism. The suggestion is based on the notion that, at the particle size relevant to exocytosis (the release of neurotransmitters at the presynaptic terminal), quantum processes will take over from thermodynamic processes. The resultant model is one that involves a mechanism of electron transfers between molecules which are triggered by stochastic state collapses of quasi-particles (Beck & Eccles, 1992). In this case the stochastic part of the mechanism is an irreducible subatomic process. Any intervention would presumably only alter the probability distribution, not necessarily make any one outcome probable ($P > 0.5$). Does this constitute a poor explanation then?

I would suggest that in some ways it does. One reason is that there are no sufficient mechanistic explanations of the stochastic behaviour of subatomic particles. By definition they are irreducible – and if the norm of explanation is that we can describe the organisation of robust entities and their activities then we have no normatively good explanation of activity at a quantum level. Therefore an explanation that *relies* on quantum level events as part of their description is not a good explanation. It may be argued that *all* explanations eventually and unavoidably reduce to the irreducible, but – intuitively – not all explanations need require a description of quantum events in order to account for the explanandum phenomenon. It does not seem useful to require complete reduction as a norm for good explanation. The alternative, it seems, is to declare a threshold of detail that must be reached. It seems to me that the only measurable outcome of expanding or contracting the scope of detail of an explanation is the change in the probability distribution of the outcomes of a

mechanism for given variable value sets of its entities. More detail, that is, more variables accounted for in the explanation, will increase the probability that – given a set of variable values – a discrete outcome will be observed. If the motivation for explanation is the desire to intervene with the entities of a mechanism in order to affect the outcome, then it would seem desirable that the outcome of an intervention should be predictable with some degree of confidence.

In neurotransmitter release, being able to intervene upon a variable to alter the observable occurrences of neurotransmitter release from one-in-five to almost zero-in-five would certainly indicate some causal relevance of that variable. However, the description of a mechanism that only allows intervention for prevention of the explanandum phenomenon does not necessarily a good explanation. The knowledge that action potentials can be reduced to a zero-in-five occurrence rate by severing the head from the body does little other than to suggest background conditions necessary for the explanandum phenomenon, but falls far short of sufficiently explaining them for the purposes of intervention.

A CAUSE VS THE CAUSE

Ultimately, the problem with the assertion that “causes need not make their effects more probable in order to explain them” stems from the ambiguity of the terms “a cause” and “the cause”. When Craver cites syphilis as “the cause” of paresis, despite only one in five cases of syphilis terminating in paresis, he is failing to acknowledge a number of underlying assumptions ((Craver, 2007, p. 26), from an original example by (Scriven, 1959, p. 130)). Certainly paresis can only occur subsequent to syphilis infection. This should not be considered a complete explanation, however, since it does not describe the mechanism by which the occurrence of paresis is exhibited. A good explanation would be a long, complex description including a full account of what constitutes a case of paresis and how this is causally linked with the phenomenon of syphilis infection. It may be the case that the development of paresis is eventually reducible to a stochastic process similar to the mutation of cancerous cells in the development of lung cancer. A description of such an event will presumably be necessary to explain lung cancer. Furthermore, demonstrating that smoking increases the risk of lung cancer is not sufficient to explain how smoking causes lung cancer. In fact, a recent study has suggested that smoking may increase the chances of lung cancer development because a chemical in cigarette smoke damages the FANCD2 protein in the epithelial cells of

the lung. This protein is one of a group inherited in a condition known as *Fanconi anemia*, and is involved in the destruction of pre-cancerous cells in early stages of mutation¹ (Hays, et al., 2008).

A description of this process, and how smoking affects it, would not constitute an explanation of how smoking causes lung cancer¹. It may be an important aspect of an explanation of the phenomenon of lung cancer in general, since it provides opportunities for manipulation and intervention. However, long-term exposure of the lungs to cigarette smoke is only one (albeit highly significant) example of the many background conditions of the mechanism of lung cancer development that should be included in a good explanation. Smoking is not "the cause" of lung cancer. The explanation of lung cancer in general is the description of the mechanism, including entities and activities that are causally relevant (N3). Smoking is a variable that increases the likelihood of lung cancer development *given a number of other variable value sets*. If a smoker develops lung cancer, the lung cancer was still not necessarily caused by smoking. In the language of normative explanation smoking was a causally relevant constituent of the general mechanism underlying the realisation of the token case. If anything were to be labelled "the cause" it would surely be the description of the causal links between the complete set of actual realised values of all the variables in the general mechanism and the outcome of lung cancer occurrence. In this sense, in order to explain the effect of lung cancer occurrence "the cause" *must* make lung cancer at least probable ($P > 0.5$), if not highly probable (depending on the interventional motivation of the explanation of lung cancer).

If Craver's claim merely means that "a cause", in the sense of any one variable value within a mechanism, need not in its taking on a value make any outcome in the probability distribution of the explanandum phenomenon probable, then that is indeed the case. If, however, the "cause" of an effect is considered a sufficient (but not "complete" or "objectively homogenous") account of the relevant value sets of variables – sufficient in that it is broad enough to allow for confident intervention and manipulation – then such a cause should make

¹ There are various theories dealing with the suggestion of causation by omission and prevention which I will not examine here. Personally I believe that omitted events do not have causal power; however good explanations may certainly require descriptions of the consequences of various omissions and preventions in order to fully account for the explanandum phenomenon. To avoid infinite lists of possible omissions and preventions, probabilities should again provide the guide for a threshold of detail of the qualified good explanation.

any chosen effect probable ($P > 0.5$) if it is to be considered to have provided a "good" explanation.

If, as Craver suggests, explanation is motivated by the intention of intervention then the outcomes of interventions must be necessarily predictable (at least above some threshold of probability). An incomplete or poor explanation may provide information about necessary conditions for the exhibition of an explanandum phenomenon. In medical biology it may be useful to know that without precondition X outcome Y will not occur (excepting some unpredictable subatomic stochastic intervention), but it is of minimal use to know that given condition X then outcome Y will occur with a 10 – 20 percent likelihood. It does not seem clear that the ability to prevent an outcome with relative confidence should be any more desirable than the ability to produce an outcome with relative confidence. In fact, in neurobiology it can potentially be dangerous to attempt interventions (for instance by means of neurochemical drugs) when the component being manipulated is considered subject to "stochastic" behaviour.

CONCLUSION

According to Craver, a good explanation in neuroscience must fully account for the explanandum phenomenon, including precipitating conditions, manifestations, inhibitory conditions, modulating conditions, non-standard conditions and by-products. It would appear that any discrete phenomenon exists as one distinct manifestation from a probability distribution of possible outcomes given a range of measurable precipitating conditions. The description of a mechanism underlying the exhibition of one discrete phenomenon will also describe the other possible outcomes in a probability distribution given the range of possible values of variables within the explanation.

If explanations in neuroscience are motivated by a desire to bring the central nervous system under control, as Craver believes they are, then it would seem appropriate that a sufficiently good explanation should describe components of a mechanism that are subject to mutual manipulability on ideal interventions with *reasonably predictable results*. That is, causes (in the sense of certain sets of values taken on by the variables of the mechanism) should be seen to make certain phenomena within the probability distribution of outcomes the most likely to occur.

It is possible, and indeed it will be common, for an explanation to fail to effectively predict outcomes by demonstrating that a cause will make an effect probable. This will occur when outcomes appear stochastic in probability. At a macrophysical level, stochastic effects will occur either when insufficient variables are included in the calculation of the probability distribution of the outcome or when one or more variables in the calculation are representative of an irreducible microphysical quantum level event. In either case, however, I see no reason to consider that such an explanation has fully accounted for the explanandum phenomenon or sufficiently achieves the desired capacity of manipulation (e.g. bringing the central nervous system under control).

The University of Adelaide

BIBLIOGRAPHY

Beck, F., & Eccles, J. (1992). Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences of the USA* 89 (pp. 11357-11361). National Academy of Sciences.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in the History of Philosophy of Science C (Biological and Biomedical Sciences)* , 36, 397-420.

Craver, C. F. (2007). *Explaining The Brain*. New York: Oxford University Press.

Hays, L., Zodrow, D., Yates, J., Deffebach, M., Jacoby, D., Olson, S., et al. (2008). Cigarette smoke induces genetic instability in airway epithelial cells by suppressing FANCD2 expression. *British Journal of Cancer* , 98 (10), 1653-1661.

Hempel, C. G. (1966). Laws and Their Role in Scientific Explanation. *Philosophy of Natural Science* , 47-67.

Hodgkin, A., & Huxley, A. (1952). A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. *Journal of Physiology* , 117, 500-544.

Ben Crisp 'Do Causes Need to Make Their Effects Probable in Order to Explain Them? The Tension Between N1 and E5 in Craver's Mechanistic Model of Explanation

Salmon, W. (1989). Four Decades of Scientific Explanation. (P. Kitcher, & W. Salmon, Eds.) *Scientific Explanation, Minnesota Studies in the Philosophy of Science* , 61, 3-219.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. (1971). Statistical Explanation. (W. Salmon, Ed.) *Statistical Explanation and Statistical Relevance* , 29-87.

Scriven, M. (1959). Explanation and Prediction in Evolutionary Theory. *Science* , 130, 130.

U.S. Department of Health and Human Services. (1982). *The Health Consequences of Smoking: Cancer. A Report of the Surgeon General*. Rockville (MD): U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health.

Woodward, J. (2003, May 9). *Scientific Explanation*. Retrieved June 9, 2008, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/scientific-explanation/>