Explanatory Pervasion and the Unity of Science

Andrew Miller
Wartburg College

The reductionist program arose out of the logical positivist era, when philosophers of science judged all other sciences against the natural sciences, which were believed to be ideal models of what science should be. Taking this idea one step further, reductionists held that all other sciences were, at least theoretically, *unnecessary*. The basic premise of reductionism is that all the laws of science can be explained, and thus rendered superfluous, by physics (Garfinkel 443). Underlying this assertion (often implicitly) is a belief in the symmetry of explanation and prediction; thus, standard reductionism believes that all the laws of science can be *derived* from the laws of physics.

The reductionist thesis no longer stands as pure and unassailed as it did during the heyday of the logical positivists. Many authors have uncovered fatal flaws in the reductionists' views. Some philosophers take their arguments a step further than merely destroying the idea of the unity of laws, though: they attempt to cast doubt upon the unity of science itself, asserting that the specialized scientific disciplines are necessarily autonomous from one another. This view is, I believe, too harsh. I intend to argue that while the sciences may not be unified by law, there is every reason to believe that they are unified by *explanation*.

Explanatory Pervasion

Due largely to the notion of what constitutes a law, reductionism can be shown to be untenable.[1] But what of a reduction of sorts in the opposite direction? What if, instead

Miller is a junior mathematics and physics major at Wartburg College. In May of 1996, he presented this paper to the Augustana Undergraduate Philosophy Conference and won Honorable Mention for Best Presentation.

of *predicting* a particular event solely from the laws of physics, I wish merely to *explain* it? Do such explanations exist?

At least two authors do not think so. Alan Garfinkel and David Owens argue against what David Owens defines as *explanatory pervasion*:

> If the occurrence of an S_1-event explains the occurrence of an S_2-event then there are physical predicates P_1 and P_2 such that (a) the occurrence of a P_1-event is sufficient for the occurrence of an S_1-event, (b) the occurrence of a P_2-event is sufficient for the occurrence of an S_2-event, and (c) P_1-events causally explain P_2-events. (Owens 65)

This is *not* identical to reductionism: for explanatory pervasion to be true, P_1 and P_2 need not be related by law. All that is required is that we can explain the higher-level events in virtue of the physical mechanisms that connect them.

Garfinkel posits that this is not possible. He illustrates his argument with the use of an example from ecology: a system of fluctuating populations of foxes and rabbits. Given that $X(t)$ denotes the number of foxes at time t and $Y(t)$ denotes the number of rabbits at time t, we can
describe the behavior of such a system with a pair of differential equations[2]:

$$\frac{dX}{dt} = aXY - bX, \qquad \frac{dY}{dt} = pY - qXY$$

Using these equations we can put forth some basic explanations for phenomena in the system. For example, if the fox population is high, the rabbit population will most likely be decreasing. Thus, if a particular rabbit gets caught and eaten, we can say that "[t]he cause of the death of the rabbit was that the fox population was high" (Garfinkel 446). According to Garfinkel, this is the *macroexplanation* of the rabbit's death, for it appeals to high-level scientific laws. We can attempt a *microexplanation*, of the event as well; Garfinkel suggests "[r]abbit *r* was eaten because he passed through the capture space of fox *f*" (Garfinkel 447).

He goes on to argue that these explanations are not equivalent, because their objects are not the same: the macro-object is the death of the rabbit, whereas the micro-object is the death of the rabbit at the hands of fox *f*, at place *p*, at time *t*, etc. Thus, this microexplanation is inadequate in that it is hyperspecific — it fails to account for small perturbations in the initial conditions. One can suppose that had the rabbit escaped death at the hands of fox *f*, he probably would have suffered the fate at the hands of some other fox, assuming the fox population remained high (Garfinkel 447). But our microexplanation does not tell us this. Neither, then, does it tell us how the rabbit might have avoided being caught, and, as Garfinkel points out, an explanation is inadequate if it does not lend itself to use for prediction and prevention (Garfinkel 448).

From these assertions, Garfinkel sets forth an alternate conception of explanation. He associates with each event a sort of phase space of initial conditions, where, small changes in the initial conditions do not result in major changes in the qualitative results except at certain critical points.

> The crucial thing we want to know is how this set of critical points is embedded in the substratum space, for that will tell us what is really relevant and what is not. Therefore, what is necessary for a true explanation is an account of how the underlying space is partitioned into basins of irrelevant differences, separated by ridge lines of critical points. (Garfinkel 452)

Is this impression of explanation really all that different from the standard account? To answer this question, we need look only so far as the boundary lines in Garfinkel's explanation space. Now, there are two ways of interpreting these boundary lines: 1) we can explain, in lower terms, *why* these boundary lines determine the changes in the system, or 2) these critical points are an inexplicable property of the system itself. If the first interpretation is true, though, we really don't have a new definition of explanation at all: Garfinkel's idea collapses into a more standard conception of explanation, in which higher-level laws can be explained, if not derived, using more basic laws. So Garfinkel's use of his phase-space model of explanation to support the disunity of science hinges upon whether these critical points are unaccountable features of high-level systems. I assert that this is not so, for reasons which I shall now explore.

The untenable nature of Garfinkel's arguments becomes clearer when we consider an idea which led him to the phase-space model of explanation: redundant causality. Garfinkel feels that in systems such as the fox-rabbit ecology, citing a particular physical mechanism of the resulting event (e.g., citing how the rabbit died) is irrelevant because

> [s]ystems which exhibit redundant causality ...
>
> have, for every consequent Q, a bundle of
>
> antecedents (P_i) such that:
>
> 1. If any one of the P_i is true, so will be Q.
>
> 2. If one P_i should not be the case, some other
>
> will. (Garfinkel 448)

Now, I don't take issue with (1) - it is just a reformulation of what David Owens calls the *multiple realization point*, the idea that each non-physical state can have associated with it many subvenient physical states. I must, however, disagree with (2). This aspect states that Q is in some way inevitable — no matter what happens, the universe will contrive to assure that some physical mechanism sufficient for Q will take place. To me, this is excessively speculative; it ascribes to the universe more order (indeed, almost an intelligence) than it could reasonably have. Furthermore, it flies in the face of fact. It is easily conceivable (though perhaps unlikely) that our rabbit could have escaped all of the foxes and gone on to live a long and happy life; his fate was never a foregone conclusion.

This whole issue can be resolved if one realizes that Garfinkel's redundant causality is nothing more than a thinly veiled theory of probabilistic laws. Based on the high fox population, the rabbit's death was highly probable, but not guaranteed. Hence, we could say 'at any given time *t*, the probability that the rabbit will be caught by a fox is high.' This probabilistic formulation captures all of the important features of Garfinkel's redundant causality: (1) is assured because we know that every time *t* will have associated with it a certain physical setup, and among those setups will be many that are sufficient (but not necessary) for the death of the rabbit; (2) is covered, and improved, by the fact that it will be highly improbable, but still possible, that the rabbit should survive for a long period of time.

Now that we recognize redundant causality as a probabilistic system, we can attempt an appropriate microexplanation of the fox-rabbit ecology. To begin, we calculate the probability of an individual rabbit *r*'s being caught and eaten in a given period of time *t* (denoted by $P(rEt)$). The differential equations tell us that in a given unit of time, the number of rabbits that are eaten is $-qXY$. The probability of a single rabbit's being eaten is that number divided by the total number of rabbits at that time, that is:

$$P(rEt) = \frac{qXY}{Y} = qX$$

From this, we see that the probability of an individual rabbit's being eaten depends only on the number of foxes in the system!

We can explain this by taking the original microexplanation (that the rabbit was eaten because he entered the fox's capture space) as a starting point and asking why this mechanism was likely to occur, i.e., asking why it was nomologically expectable that the rabbit should enter a fox's capture space. It is reasonable to suppose that the size of an individual fox's capture space (relative to rabbits) is a function of both the rabbit's and the fox's attributes (their individual speeds, reaction times, senses of smell, etc.). So we let $S(A_R, A_F)$ denote the size of this capture space, where A_R and A_F are the relevant attributes of rabbits and foxes. Then the total area covered by the capture spaces of all the foxes at any time t is less than or equal to $X \quad S(A_R, A_F)$ ("less than or equal to" because of the possibility of separate foxes having overlapping capture spaces). Assuming that the foxes and rabbits are confined to a region with area R, we can say that the probability that a rabbit will be in the capture space of a fox, and hence the probability of his being eaten, is given by

$$P(rEt) \leq \frac{X \cdot S\,(A_R, A_F)}{R}$$

which is, as we predicted, a function of the number of foxes in the system.

And so we have a formula, stated in terms of probabilities, spatio-temporal coordinates, and the biological attributes of foxes and rabbits, that is a potential explanation of the behavior at the sociological level. The adequacy of this explanation can be tested by applying the conditions stated earlier by Garfinkel - namely, that an adequate explanation must lend itself to prediction and prevention. Clearly, with this model I am able to predict how likely it is for any one rabbit to die, and from there to derive the overall results at the sociological level. This model also shows me what I must do if I wish to prevent the death of rabbits (or at least slow down the population decline). I can do one of three things: decrease the number of foxes; decrease the size of the foxes' capture spaces by improving my rabbits or somehow handicapping my foxes; or increase the region in which the foxes and rabbits are confined. Of these three, only one is demonstrated in the original differential equations - the other two are "hidden" in the constant q. It appears that I have a successful microexplanation that actually has more descriptive content than the macro-explanation!

So we see that Garfinkel's arguments are insufficient to show the absence of explanatory links between the higher-level scientific laws and physics. The case is far from closed, however. The systems considered so far by Garfinkel and myself are a good deal less complicated than many systems which arise in the higher-level sciences. Whereas the examples so far have been ones in which the physical realizations of the higher-level events have been largely homogeneous, there are many cases in which the physical mechanisms are quite different from one another. In these cases, one would need to formulate a separate explanation for each type of mechanism.

David Owens makes this point when he considers (in a more direct fashion than Garfinkel) the thesis of explanatory pervasion in light of an economic example: a monetary exchange. Owens looks at Fisher's Law, which states that if there is an increase in the money supply in an economic system, then there will be an increase in prices (Owens 62). Now, increases in the money supply and increases in prices can be realized in many ways. The discovery of a new gold mine, the exploitation of a new industry, the lowering of interest rates, and so forth are examples of an increase in money supply. Similarly, raises in prices can be reflected in the real estate market, the agricultural industry, the entertainment industry, and every other market in the economic system. Thus, the P_1 and P_2 in the explanatory pervasion thesis shall have to be broadened to allow us to apply them to heterogeneous sets of physical mechanisms. The multiple physical realizations of these economic predicates shows us that a necessary condition for the truth of explanatory pervasion is *agglomerativity*: "Agglomerativity of Causal Explanation: If A causally explains B and C causally explains D, then A&C causally explain B&D" (Neander et al. 460). Furthermore, since there will be no direct explanatory linkage between the physical realizations of an increase in the money supply and an increase in prices, we will have to instantiate a chain of explanations to connect the two. On the basis of this argument, one can conclude that an additional necessary condition for explanatory pervasion is *transitivity*: "Transitivity of Causal Explanation: If A causally explains B and B causally explains C, then A causally explains C" (Neander et al. 460). Owens proceeds to attempt to show that both agglomerativity and transitivity do not always apply to explanation, casting doubt upon the thesis of explanatory pervasion.

First, we examine agglomerativity. To attack this thesis, Owens gives the following Aristotelian example: "A man eats spicy food, and gets thirsty, and so goes out to the well, where he meets some ruffians who happen to be passing and who kill him. They did not come for the purpose of finding him, nor did they lure him there" (Owens 72). Here we have a consequent (the man's murder) which is the result of joining two component physical processes (the arrival of the ruffians at the well at time *t*, and the arrival of the man at the well at time *t*) each of which have causal explanations: the man's eating spicy food, and the ruffians wishing to rest at the well. However, it cannot be said that the joining of these explanations causally explains the joint arrival of the man and the ruffians at the well. This joint arrival has no proper explanation; it is a coincidence.

Owens claims that the reason that agglomerativity fails in this case is that the individual explanations have no common element, just as the explanation of how lowering interest rates causes inflation has no (or very few) elements in common with an explanation of how the discovery of a new gold mine accomplishes the same effect (Owens 72). He contrasts these examples with a situation in which agglomerativity *does* hold because of such a common element: the joint arrival of all of the members of an orchestra at an orchestra hall for a rehearsal can be explained by the fact that each member heard the conductor announce the time of the rehearsal (Owens 72).

Karen Neander and Peter Menzies agree with Owens that the agglomerativity of causal explanations does not always hold, but they disagree with his arguments to that effect. To illustrate their point, they redescribe the Aristotelian example with different constituent explanations:

> The victim left home at 6 a.m., walked a distance *D* at an average velocity of *V*, and consequently arrived at the well at noon. The ruffians decamped at 9 a.m., walked a distance of *D*/2 at an average velocity of *V*, and also arrived at the well at noon. (Neander et al. 461)

On the basis of these explanations, the simultaneous arrival of the man and the ruffians ceases to be coincidental; it is predictable from our explanations. But these explanations share no common element, for surely the man's motion is independent of any causal factors that brought about the ruffians' motion. Hence, Owens's arguments are deficient.

As an alternative reason to reject agglomerativity, the authors develop the idea that explanations are often invoked relative to a *contrast class*[3]:

> If we schematize the simplest kind of event as an object *a*'s having property *F* at time *t*, then a request for a causal explanation of this event may be calling for different things: it may be calling for an explanation of why the object *a*, rather than some other object, has *F* at *t*; or it may be calling for an explanation of why *a* has property *F*, rather than some other property, at *t*; or it may be requesting an explanation for why *a* has *F* at time *t*, rather than at some other time. (Neander et al. 462)

Relative to this idea, we can illuminate the differences between these two descriptions of the Aristotelian example. The most important factor of the event that we are considering is that the ruffians and the man arrived at the well at the same time. So we seek an explanation which will tell us why they all arrived at the well at that particular time, rather than some other time. In Owens's explanation, the component explanations tell us nothing about why the ruffians and the man arrived at that specific time; they merely tell us why they went to the well, as opposed to going somewhere else. Neander and Menzies's explanations, on the other hand, give a precise explanation of the times of arrival, and joining them does explain the two parties' simultaneous arrival. Accordingly, the authors conclude, "as far as we can see, agglomerativity holds provided that the contrast class of the compound explanation matches the contrast classes of the individual explanations" (Neander et al. 463).

On this account, we can see that the explanations of the individual instantiations of Fisher's Law still do not agglomerate: just as citing the specific mechanism for the rabbit's death did not explain *why* he was caught, citing the mechanisms through which the rise in prices is achieved is insufficient to show why this inflation took place. We are not asking why the prices rose at that time, in that place, etc., but rather why they rose at all. If we could formulate some explanation that accounted for the reasons behind the inflation, as I did with the foxes and rabbits, I see no reason to believe that agglomerativity would not hold for those explanations.

To challenge the idea of the *transitivity* of causal explanations, Owens looks at a well-known nursery rhyme:

> For want of a nail the shoe was lost,
>
> For want of a shoe the horse was lost,
>
> For want of a horse, the rider was lost,
>
> For want of a rider the battle was lost,
>
> For want of a battle the kingdom was lost,
>
> And all for want of a horseshoe nail. (Owens 75)

Owens casts doubt upon the validity of the inference in the last line. Neander and Menzies again agree with Owens that the transitivity of explanation does not always hold, and again they make strong claims against the efficacy of Owens's arguments. Since the details of Owens's position are neither essential nor illuminative of the issue at hand, I will consider only the stance of Neander and Menzies on the issue.

The authors invoke another facet of explanation to support their assertion: explanatory relevance. They note that "in asking for the causal explanation of some event ... one is asking for an explanatorily *relevant* causal condition" (Neander et al. 464). Thus, the question arises: Are we justified in saying that P_1-events of the type described in the thesis of explanatory pervasion are explanatorily relevant to P_2-events?

Neander and Menzies contend, correctly, in my opinion, that the answer is no. In most cases, the explanatory chain will be too long for this to be true: "transitivity will fail in those cases in which the probability of the explanandum event falls below [a] required threshold when we trace its causal ancestry far enough back" (Neander et al. 465). The nursery rhyme is one such case. The loss of a battle could reasonably bring about the loss of a kingdom; and one could suppose that the loss of a particular rider would be sufficient for the loss of the kingdom as well, *if* that rider were the king or some other vitally important political figure. But that is as far as we may go, say Neander and Menzies, for the loss of a horse does not make the loss of the kingdom sufficiently probable for us to say that it is explanatorily relevant.

Is this true? Surely, one would think that if in reality the battle was lost for lack of one rider; and that rider was lost only bcause his horse collapsed, and that this in turn happened only because of a loose shoe, then the missing nail would indeed be necessary in an explanation of the event. To resolve this conflict, one must distinguish between *explanatory* relevance and *causal* relevance. The missing nail is causally relevant to the loss of the kingdom in the same way that "rabbit *r* entered into the capture space of fox *f* at time *t*" is causally relevant to the death of the rabbit. But it is not explanatorily relevant, for the same reason that the rabbit's movements were not explanatorily relevant to its death: an explanation based on the missing nail is not useful for prediction and prevention. Had a blacksmith taken more time to ensure that the horse's shoe was firmly affixed, would that have prevented the loss of the kingdom? Common sense tells us no: if the kingdom were so fragile that the death of one rider would determine its fate, we would guess that some other minor factor (a rusty sword, for example) would have caused its downfall. There are other, more important, factors in the kingdom's history (its military strength, recent political happenings, economic stability, etc.) that acted to bring about this fragility, and a true, useful, explanation would cite these factors.

It is interesting to note that this explication of the transitivity of explanation is very closely related to the vindication of agglomerativity. Again, we are dealing with contrst classes. The missing nail answers the question, "Why did the kingdom fall at *this* time, in *this* way, in *thise* as opposed to falling tomorrow, or to a different country, etc.?" An explanation of the kingdom's collapse would more likely be seeking hte answer to the question "Why did the kingdom fall *at all*, as opposed to continuing on for fity prosperous years?" Thus, a particular physical realization of the kingdom's downfall is not explanatorily relevant.

Now that we see that explanatory trnasitivity is ot true in all cases, we are forced to reject the original concept of explanatory pervasion inasmuch as it is too strong. All is not lost, though, for we may still hold on to agglomerativity as long as we pay adequate attention to the contrast classes of high-level explanations. Thus, we can formulate a weaker version of thethesis. Neander and Menzies call it *explanatory pervasion*:

> If the occurrence of an S_1-event explains the occurrence of an S_2-event then there are physical events P_1 and P_2 such that (a) the ocurrence of the P_1-event is sufficient for the occurrence of the S_1-event; (b) the occurrence fo the P_2-event is sufficient for the occurrence of the S_2-event; and (c) there is *a series of explanations* linking the p_1-event with the P_2-event. (Neander et al. 466)

Conclusion

Now we see that those authors who would reject a thesis of explanatory pervasion are being overly critical of the positivists' views. Indeed, a belief in the disunity of science and the autonomy of the special sciences goes against both commonsense and common practice in the scientific community. This does not mean that traditional formulations of explanatory pervasion emerge from this debate unscathed. On the contrary, as I have shown above, in light of the failure of the transitivity of explanation it is necessary to reformulate the thesis in the manner of Neander and Menzies. This new conception of explanatory unification, as opposed to theory unification, provides a promising common ground for philosophers and scientists alike to share information across disciplines.

Works Cited

Boyd, Richard, Philip Gasper, and J.D. Trout, eds. *The Philosophy of Science*. Cambridge: The MIT Press, 1991.

Carnap, Rudolf. "Logical Foundations of the Unity of Science," Boyd et al. 393-404.

Fodor, Jerry. "Special Sciences," Boyd et al. 429-41.

Foss, Jeffrey. "Materialism, Reductionism, Replacement, and the Place of Consciousness in Science," *The Journal of Philosophy* 92 (1995): 401-29.

Garfinkel, Alan. "Reductionism," Boyd et al. 443-59.

Neander, Karen and Peter Menzies. "David Owens on Levels of Explanation," *Mind* 99 (1990): 459-66.

Oppenheim, Paul and Hilary Putnam. "Unity of Science as a Working Hypothesis," Boyd et al. 405-27.

Owens, David. "Levels of Explanation," *Mind* 98 (1989): 59-79.

Popper, Sir Karl. "Scientific Reduction and the Essential Incompleteness of All Science," *The Open Universe: An Argument for Indeterminism*. Totowa: Bowman and Littlefield, 1982.

— —. "Further Remarks on Reduction, 1981." *The Open*

Universe: An Argument for Indeterminism. Totowa: Rowman and Littlefield, 1982.

Trout, J.D. "Reductionism and the Unity of Science: Introductory Essay," Boyd et al. 387-92.

NOTES

1. Jerry Fodor, in "Special Sciences", does an especially good job of arguing against reductionism on appeals to ideas of lawlikeness.

2 . In these equations, the product *XY* represents the number of encounters between foxes and rabbits; *a*, *b*, *p*, and *q* are constants whose values are determined by fitting the observed data in an actual system. These equations are called the Lotka-Volterra equations, and readers who wish to explore their derivation in more detail are referred to Garfinkel's sources: Braun, *Differential Equations and Their Applications* (New York: Springer-Verlag, 1975) and E.C. Pielou, *An Introduction to Mathematical Ecology* (New York: Wiley-Interscience, 1977).

3. This idea was, I think, implicit in Garfinkel's arguments. I believe his reasoning for why microexplanations often seem to miss their mark lines up perfectly with this idea.