Fitness

Alexander Rosenberg

The Journal of Philosophy, Vol. 80, No. 8. (Aug., 1983), pp. 457-473.

Stable URL:

http://links.jstor.org/sici?sici=0022-362X%28198308%2980%3A8%3C457%3AF%3E2.0.CO%3B2-T

The Journal of Philosophy is currently published by Journal of Philosophy, Inc..

mixtures, and to a relation between the being part of relation and homeomerism. Specifically, there are certain kinds of stuff A, quantities of which contain quantities of other kinds of stuff B as part if and only if the B exists homeomerously in the A. An example is the water in the air, which is part of the air if and only if it is in it homeomerously, but otherwise that water is merely in the air like a bird, without being part of the air.

This I find quite exciting.

RICHARD SHARVY

The Avondale Institute

FITNESS*

D EBATES about the cognitive status of the Darwinian theory of natural selection should have ended long ago. Their persistence reflects the steady failure of biologists and philosophers of science to treat the notion of fitness as the quite ordinary theoretical term which in fact it is. Even the rare expositions of fitness and its role in evolutionary theory that have been correct have failed to put the methodological controversy over this theory to rest.[1] In this paper I shall show that 'fitness' differs from an ordinary theoretical term, like temperature, not in kind, but only in degree; that this difference sets limits on the measurement of fitness; that these limits give the theory of evolution its undeserved reputation for vacuousness. I then apply these conclusions about fitness to laboratory experiments in evolution, with a surprising re-

[1] For example, D. Hull, *The Philosophy of Biological Science*, and M. Ruse, *The Philosophy of Biology*, both address the allegation that the theory of natural selection is vacuous but they fail to explain why it persistently attracts this false charge. Other recent attempts to refute the charge, such as Mary Williams, "Falsifiable Predictions of Evolutionary Theory," *Philosophy of Science*, 40 (1973): 518-537, or S. K. Mills and J. H. Beatty, "The Propensity Interpretation of Fitness," *Philosophy of Science*, 46 (1979): 263-288, neither correctly diagnose the source of this error, nor provide effective remedies for it, and they generate some further obstacles to the dissolving of the mistake. For details of these defects see, A. Rosenberg, "The Supervenience of Biological Concepts," *Philosophy of Science*, 45 (1978): 368-386, and "On the Propensity Definition of Fitness," *Philosophy of Science*, 49 (1982): 268-273.

sult for biology: in the only setting which can precisely confirm the theory of natural selection that theory's claims about fitness are superfluous. This conclusion has ramifications for intentional psychology; for the same arguments which show that fitness is otiose in the biologist's laboratory, show that intentionality cannot be expected to play an important role in the psychologist's laboratory.

<div align="center">I</div>

Fitness is a relational property, reflecting the interaction of an organism and its environment. How we apply the theory of natural selection to explain adaptation, competition and evolution, depends on how we measure the fitness, and especially the differences in fitness among organisms.[2] We need to be able to make comparisons between members of the same species in the same environment, between members of the same species in different environments, between members of different species in the same environment, and between members of different species in different environments.[3]

An organism's fitness is relative to its environment, but an environment is not a spatiotemporal region. It is a region in more than four dimensions. Its dimensions are given by the set of forces that can interact with the organism's properties to determine its fitness level.[4] Thus, two environments might be spatiotemporally identical, and yet distinct from one another; for example an open field may provide distinct overlapping environments for grouse, squirrels, insects, certain flora, etc. We identify environments along three spatial and one temporal dimension, but they lie in a space of a

---

[2] Actually, differences in fitness level may be predicated of organisms, genomes, gametes or any other types of items that constitute a biosphere, a line of descent. These items need not satisfy ordinary intuitions about their organic constitution, or methods of reproduction. Indeed they had better not be so restricted if the theory of natural selection, which specifies their mechanism of evolution—or non-evolution—is to be nomologically universal. For expository convenience I shall speak of organisms, though my claims should be understood as holding for any of the items that can have evolutionary fitness.

[3] For a discussion of some of the practical problems surrounding the measurement of fitness, and suggested means of dealing with them, see T. Prout, "The Relation Between Fitness Components and Population Prediction in Drosophila, I and II," *Genetics*, 68 (1971): 127-149, 151-167, T. Prout, "The Estimation of Fitness from Population Data," *Genetics*, 63 (1969): 949-967. It is clear that fitness is a quantitative functor, with a natural zero, although for many purposes only ordinal differences in fitness levels need to established.

[4] This fact is part of the explanation for why mathematical biological theory has increasingly borrowed the formalism and adapted the theorems of mathematical economics: theorems about the existence, stability and uniqueness of equilibria under varying conditions have been demonstrated in economics by appeal to topological properties of Hyper-spaces. These results are being increasingly adopted in ecology and evolutionary theory. See for instance, R. May, *Stability and Complexity in Model Eco-systems* (Princeton, N.J.: University Press, 1973).

much larger set of dimensions, along each of which variations affect fitness levels. Exactly how large the dimensionality of an evolutionary environment is we cannot tell, because we know too little about the structure and behavior of organisms, and about how external forces affect them. Even if we had adequate theories about the structure, behavior and environment of organisms, the practical difficulties surrounding any attempt to fully explain why an organism had a particular fitness level would be staggering. The number and complexity of such theories would make reasonably complete explanation of an organism's fitness level unmanageable in length, and incapable of actually predicting biologically interesting consequences of its fitness level.

Thus there are many different ways in which the same level of fitness may be possessed, and there is consequently no one-to-one relation between a given level of fitness and a manageable set of its causal conditions. We cannot identify levels of fitness by appeal to the properties of organisms and environments that determine it. The "one-many" relation between fitness on the one hand and the determining properties of organisms and environments on the other is that of "supervenience" in the sense expounded by Jaegwon Kim.[5]

Because of the one-many relation between fitness and its determinants, fitness must be measured *by its effects*. Now the measurement of a theoretical functor by measuring its effects is a common practice in physical science. Changes in temperature are measured by an alcohol thermometer which measures the effects of such changes on linear expansion of alcohol in a narrow tube. Linear expansion in the thermometer does not define changes in temperature, it provides a way of measuring its magnitude. Because a thermometer measures temperature changes by measuring their effects, we can explain its function by appeal to the causal relation between temperature and linear expansion. If temperature were defined in terms of linear expansion we could not do this. The operation of a thermometer is explained by citing the very phenomenon it measures. Moreover, the measurements thermometers give often need to be corrected; indeed at extreme temperatures, these instruments do not give readings at all. Beyond the boiling and freezing points of alcohol, thermometers no longer respond to changes in temperature and cannot be employed to measure them. This fact is explained by appeal to thermodynamics and chemical theories of

[5] "Supervenience and Nomological Incommensurables," *American Philosophical Quarterly*, 15 (1978): 149-156; Kim's strategy has been applied in Rosenberg, "Supervenience of Biological Concepts" note 1 above.

the ionic and vaporization properties of alcohol. Beyond the boiling and freezing points of alcohol different measuring devices must be employed, and these instruments can also be compared with the alcohol thermometer's readings at intermediate temperatures. Differences between alternative devices' readings must themselves be explained, and some measuring instruments corrected in the light of others. To determine which are the most reliable and to effect the needed corrections we employ theories about heat and about other non-thermodynamic properties of the instruments. Some of those theories may even show that it is possible to measure temperature changes by appeal to their causes instead of their effects. For instance, we can measure changes in temperature by measuring changes in electrical resistance which causes it. But of course such a possibility is impractical for many purposes; measurement in terms of subsequent effects is preferable, even when ones that rely on prior determinants are in principle available.

The measurement of fitness, as it figures in the theory of natural selection, differs from that of temperature in two respects. Fitness is a function of more complex relations among more variables than temperature is. And there is no theoretical edifice to stand behind fitness, to explain and correct measurements of it, as there is for temperature. Such theories may eventually be found, but they are not on the scientific horizon. Because of these two features of the conceptual position of fitness, measurement of it depends on its effects even more fully than the measurement of temperature does.

The fitness of an organism is measured by counting progeny, either of the organism, its ancestors and/or descendants, or some subset of the progeny in a branching tree of descent. But while these demographic counts are the units in which fitness is measured and are the only common coin of evolutionary comparisons, it should be obvious that as they stand, they are highly unsatisfactory measures of fitness. Consider for example identical twins that have reached reproductive maturity in the same environment. Presumably two such peas in a pod will have the same level of fitness. But if through some entirely random event one of them is destroyed, then the other will have vastly more off-spring; if we measure fitness by number of off-spring, without *correcting* for the intervention of short-run, non-selective forces, we will wrongly conclude that these two biologically identical organisms had different levels of fitness. The theory of natural selection assures us that in the long run, persistent differences in the level of fitness, no matter how small, between competing members of a biosphere will lead to differences in reproductive levels. That is why we employ short term dif-

ferences in reproductive success to *estimate* fitness. But short term trends may not reliably reflect long term ones and therefore uncorrected fitness-estimates may be wrong. Any actual level of fitness is consistent with any short-term level of reproduction, even zero reproduction, provided that the lack of descendent results from forces exceptional and/or indeterministic enough to be deemed random with respect to evolution (like the circumstances which kill Schrödinger's cat).

We can tell that a particular property—like mimicry, or disease resistance—makes for fitness by seeing that it contributes to optimum or satisfactory design. To recognize that a particular trait is suitable for attaining a certain end in a given environment we need not enquire into reproductive success; instead we may employ theoretical or engineering information to identify alternative strategies for meeting a design-requirement, and to grade the efficiency of those alternatives in the light of environmental and organic constraints. Pursuing this approach enables us to identify some of the determinants of fitness in particular cases, and so grounds restricted judgments of comparative fitness on its causes instead of its effects. But design criteria are often myopic. They can easily lead to misidentifying the more fit as the less, and *vice versa;* thus a parasite more efficient at wasting its host may be wrongly identified as fitter than one which survives it just because the survivor is less efficient and therefore does not destroy its host. At best, employment of design criteria may enable us to make restricted qualitative non-fungible comparisons of fitness, with respect to a small number of properties, and a restricted class of environments. Optimum design cannot play the systematic quantitative role required of a general measure of fitness that the theory requires. Fitness levels must be measured in a coin common to many biological systems if it is to play its explanatory function in the theory of natural selection.[6]

II

Fitness levels are in fact related to actual reproduction rates in a way familiar to philosophers who have reflected on problems of probability.[7] The hypothesis that an organism, in a given environment, has a given level of fitness is like the hypothesis that a given die is a fair one. This latter hypothesis is consistent with any finite sequence of outcomes of rolling the die, even 10^6 straight '2's. No finite number of outcomes of rolling the die can refute the hy-

[6] The role of optimality models in evolutionary theory is insightfully discussed in J. Beatty, "Optimum Design Models and the Strategy of Model Building in Evolutionary Biology," *Philosophy of Science.* 47 (1980): 532–561.

[7] See for example, Mills and Beatty, *op. cit.,* note 1 above.

pothesis. Still the claim that the die is fair has empirical content, which is why it can satisfactorily explain a run of rolls in which each face comes up the same number of times. This is true even though our only practical basis for believing that the die is fair is the finite number of rolls we have observed it to have been thrown in the past. There is in principle another independent way to assess the fairness of the die: by examining its material composition and structure. But although this method holds out the advantage of being independent of actually rolling the die, it has the disadvantage of being utterly impractical for any one who might be seriously interested in the hypothesis. If there were no possible way to determine whether a die was fair without actually rolling it, then the fact that one can roll any finite sequence whatever, would deprive the fairness hypothesis of all its explanatory power with respect to actual sequences. The same must be true of fitness. If fitness levels could not even in principle be measured by anything but levels of reproduction, then differences in fitness could not explain differences in rates of reproduction, and so could not explain evolution. However, the fact that biologists do correct values for fitness given by actual shortrun reproductive rates, shows that they do have access to alternative means; for example they can appeal to optimal design for correcting comparative judgments in particular cases. The trouble is that without general theories of organism structure, function, and environmental dependence, these corrections do not have the systematic basis that corrections of thermometers do , or even the basis that the evaluation of the fairness of a die in principle has. And so these corrections can seem *ad hoc.* They are not in reality *ad hoc,* nor are they merely designed to preserve the theory from falsification. But they are based on practical and theoretical considerations that are heterogeneous, restricted in their application, and too piecemeal in their bearing on fitness to find a place in any canonical statement of the theory of natural selection.[8]

[8] The parallel between the hypothesis that a die is fair and the attribution of a given level of fitness is so strong that some have been tempted to treat fitness as an objective propensity on a par with objective probabilities or propensities in nature. The philosophical problem of connecting a long-run relative frequency or a logical probability to finite runs is sometimes solved by simply asserting that the subject of the probability statement simply has an objective probability of behaving in a given way, a probability not analyzable into epistemic or evidential terms. This solution to the interpretation problems of probability seems to have all the advantages of theft over honest toil, for it does not solve the philosopher's problem, it shelves it by positing an otherwise unaccountable dispositional property, one which in the nature of things has no basis on occurrent properties of the object. Defining fitness as an objective propensity to reproduce, as R. Burian does in "Adaption," in M.

At least one cause should now be clear for why the theory of natural selection is persistently charged with vacuity and circularity, and denied the cognitive status of other scientific theories. Such charges are natural given the features of the theory and its key theoretical term. If fitness were defined in terms of differential reproduction then the theory would be guilty as charged. But if differential reproduction is a contingent, causal consequence of differences in fitness, and if it also provides the only practicable general means of measuring fitness, then anyone who supposes that measurements give the meaning of theoretical terms can be excused for mistakenly concluding that the theory is a grand tautology. This mistake is characteristic of operationalism, a doctrine still widely embraced among natural and social scientists. But even those not imbued with this discredited view may draw the same conclusion. Measurements of short term rates are explained away, or corrected, in a manner that is theoretically piecemeal; piecemeal just because of the biologist's ignorance of the general theories that can systematize his corrective tactics. But if piecemeal corrections are mistaken for *ad hoc* steps taken simply to preserve the theory, it is easy to infer that the theory is unfalsifiable and vacuous. Thus the operationalist deprives the theory of natural selection of any explanatory force by assimilating the key term of the explanans, 'fitness' to the crucial terms of the explanandum, 'differential reproduction'. On the other hand those who claim that the theory is unfalsifiable infer from the current want of an independent general theory for correcting demographic estimates of fitness, that no such theory is possible at all, and that biologists are therefore not entitled to make the piecemeal adjustments to fitness estimates required in the light of the theory. These two mistakes are often jumbled together. They are likely to be persistent as well, for there is no prospect of finding theories that will provide manageable alternatives to measuring fitness in terms of its effects on reproduction, its explanadum phenomenon.

Fitness is a primitive or undefined term *with respect to* the the-

---

Greene, ed., *Dimensions of Darwinism* (New York: Cambridge, forthcoming) or as Mills and Beatty do, "The Propensity Definition" note 1 above, has all the metaphysical and epistemological defects of the notion of objective probability, and some of its own besides. Moreover, as a solution to the problem of explaining and dispelling the vacuity of the theory of natural selection, it is ultimately unavailing See Rosenberg, "On the Propensity Definition of Fitness" Note 1 above. The main problem with propensity definitions of fitness is that they sever the direct connection between reproduction and fitness by interpolating a propensity which cannot be explained within the theory, and so regenerate all the problems that beset fitness itself within the theory. Dispelling these problems without recourse to objective propensities or probabilities shows the superfluousness of this notion.

ory of natural selection.[9] This is a crucial fact about the theory that
few have recognized.[10] One cause of this failure to recognize that
the term is a theoretical primitive is that 'fitness' has a meaning in
ordinary language (like 'force'), and informal expositions of the
theory of natural selection capitalize on this meaning. When fitness
is explained or illustrated, by piecemeal non-evolutionary means, it
is not recognized that the interpretation is not part of the theory of
natural selection, but rests on biological theory and observation
that is independent of evolutionary theory. Thus, according greater
fitness to one organism than another on the basis of design criteria
is useful in limited respects, but is easily misrepresented as reflect-
ing an implicit definition that interprets the term *within* the ambit
of evolutionary theory. Because no general interpretation of 'fit-
ness' can be given within the theory, evolutionary biologists who
feel constrained to define their central theoretical term sometimes
incautiously state that fitness is differential reproduction, thereby
turning the theory into a tautology.[11] Their error is to suppose that
every term in a theory is defined within that theory. Since the only
feature fitness has that could be employed in a general characteriza-
tion of it is its effects on reproduction, biologists are apt to adopt
this devastating definition wherever their exposition of the theory
seems to require definitions of its key terms. To avoid this tempta-
tion all that needs to be seen is that while 'fitness' is primitive with
respect to one theory, it can be interpreted in other theories or even
by non-theoretical, apparently *ad hoc* considerations about the
items of which it is predicated.

Of course these interpretations will not take the form of univer-
sal general statements about fitness that are both true and managea-

[9] Of course since no scientific theory has a unique axiomatization, none contains
a term that is undefined in all its possible axiomatizations. Which axiomatic presen-
tation, and which partition of primitive and defined terms we prefer is a matter of
non-formal considerations. In asserting that 'fitness' is a primitive term I do not
mean to deny that there are logically possible axiomatizations of the theory in
which it is defined, and not primitive. Rather I claim that in the axiomatization
which is most adequate or convenient for expressing and applying the leading ideas
of the theory of natural selection 'fitness' is a primitive term. And further, when this
axiomatic system is given its intended interpretation the features "semantically rele-
vant" for fitness are not to be found in the claims of the theory about fitness. For an
account of the notion of "semantic relevance" see P. Achinstein, *Concepts of Science*
(Baltimore: Johns Hopkins Press, 1968) ch. I-III.

[10] This fact about fitness is most clearly illustrated in M. B. Williams, "Deducing
the Consequences of Evolution," *Journal of Theoretical Biology*, 29 (1970): 348-385.
This important paper provides the best account available of the axiomatic structure
of evolutionary theory.

[11] See for instance E. O. Wilson, *Sociobiology* (Cambridge, Mass.: Harvard, 1975)
V. Grant, *Organismic Evolution* (San Francisco: W. H. Freeman, 1977); T. Dobzhan-
ski, *Genetics of The Evolutionary Process* (New York: Columbia, 1970).

bly short. Because of the number and the heterogeneity of causal forces that determine fitness, and because of our ignorance of the general theories that describe them and their interactions, we cannot expect an interpretation of 'fitness' that is anything like the interpretation of temperature, as mean kinetic energy, involved in the reduction of the ideal gas law to the kinetic theory. No doubt there is a vast though as yet unknown collection of non-evolutionary theories that together could generate a baroque and useless disjunctive interpretation of fitness, but this is a reflection of the fact that with respect to the theory of natural selection, this term is a primitive. This semantic vacuum is filled by misguided attempts to define fitness.

III

Temperature is a property simple and accessable enough to be entrenched in scientific theories that are now two hundred years old and more. Once temperature was connected to the mean kinetic energy of the molecules of a gas, thermodynamics was freed from theoretical dependence on such temperature measuring devices as the skin and the alcohol thermometer. Indeed the situation was reversed, and the operation of at least some of these measuring devices became explicable. More important, the theory could be employed together with temperature measurements to make predictions about how a system responds to thermodynamic disturbances of varying kinds. And of course where the theory *cum* measurement was disconfirmed, the measurement was more often than not rejected or corrected. By contrast Darwin could make no particular predictions or for that matter retrodictions on the basis of evolutionary theory alone, or at least none *specific enough* to confirm the theory in the opinion of those not already wedded to it.[12] The importance of Darwinian theory does not lie in its predictive strengths, for as we have seen, the nature of its key explanatory concept, fitness, pre-

---

[12] The persistent appeal to confirmation of the theory of natural selection by industrial melanism in the English moth, *Biston Betularia* reflects this confirmational weakness. The test of the theory provided by this phenomenon is a highly qualitative retrodiction, in which the theory can not predict the quantitative extent, but only the relative directions of the change in wing color. There are no other instances as complete or as dramatic in their confirmation of evolutionary theory, and for this reason the study is universally cited. But because of the singularity of the phenomenon, its purely qualitative bearing on the theory, the relatively short space of time during which the adaptive changes are alleged to have been manifested, the degree to which fitness in this case is identified by design considerations, and not differential reproductive rates *in vivo*, this case does not provide the kind of detailed, precise, strong confirmation the theory requires to convince skeptics. In making these points I do not mean to deny that the phenomenon does confirm the theory, only that it does so strongly.

cludes such strength. The importance of the theory lies in the freedom it provides biologists to view natural phenomena as just that, as natural, and not as the creation of an artificer with designs for natural phenomena. It was only with the advent of other nonevolutionary theories that the theory of natural selection came to generate specific predictions substantial enough to confirm it. The theory acquired its greatest increase in explanatory power when first harnessed together with Mendelian genetics. For this theory provides an independent means of identifying characteristics subject to inheritance.[13] What accounts for evolution is the fitness or lack of it these traits confer. And differences in fitness can be identified on a case-by-case basis through appeal to piecemeal theory and other unsystematic means. For example an *Escherichia coli* bacterium resistant to tetracycline is fitter in a tetracycline rich environment than one lacking this trait. So we may use non-evolutionary, design considerations to predict that one is more likely to survive and reproduce than the other. But these case by case determinations will not enable the evolutionary biologist to make many of the comparisons he wants to make. He cannot say whether a tetracycline resistent bacterium is fitter than an ampicillin resistent one in the ambient environment, when the respective environmental level of each of these drugs is unknown. To answer this question he must examine growth rates of the bacteria cultures, that is, he must measure population changes. But this involves him in the circle of evidence and explanation that the notion of fitness engenders. This circle deprives the theory of tests that employ measuring "devices" independent of the theory. Only in simple laboratory settings, involving huge numbers of very simple organisms, whose genomes are relatively small and well known, in environments subject to complete control, can the theory of evolution be applied, tested and confirmed *with all the precision* that is demanded by its critics and desired by its defenders.

But, ironically, in these laboratory settings the theory is dispensable! Everything which it can explain about what is happening in a well controlled laboratory experiment, can be explained more deeply, more directly and in greater detail by physiological and biochemical principles that *do not mention the evolutionary concept of fitness*. When enough detail, theoretical and experimental, has

[13] Mendelian theory must be distinguished from the theory of natural selection, although they are run in harness as "the synthetic theory of natural selection." The importance of distinguishing them is treated in William Provine, *The Origins of Theoretical Population Genetics* (Chicago: University Press, 1971) and A. Rosenberg, "Genetics and the Theory of Natural Selection: Synthesis or Sustenance?" *Nature and System*, 1 (1979): 3-15.

been gathered to make a prediction that specifically confirms the claims of evolution about the maximization of fitness, the theory of natural selection, and the notion of fitness become *superfluous:* They are no longer required to effect the prediction or to explain the occurrence of the predicted phenomenon. And the prediction that can be extracted from the theory in such cases is at best generic, while the explanation it provides will be qualitative at most. The theory of natural selection and any appeal to evolutionary fitness are superfluous in these settings because in them we are already able to identify and measure directly the determinants of differential reproduction, without making a detour through estimates of fitness based on these reproductive differences. Although it is superfluous in these cases, the theory is also strongly confirmed by them just because in these laboratory experiments the biological processes are so rapid, so protected from the intervention of unknown or unexpected forces, and the number of organisms so huge, that what evolutionary theory tells us will happen in the long term, happens in the short term, and happens invariably (provided the experiment is well designed). Outside the laboratory biological processes are slow, intervention of unknown forces is the rule, the number of organisms is small, and we must have recourse to fitness if we are to explain evolution at all.

The superfluousness of the theory of natural selection for explaining and prediction evolution in the laboratory is reflected in the reluctance many show to accepting laboratory experiments and simulations as tests of the theory of natural selection.[14] They are of course tests of it, but tests in which the predicted events can be explained in greater detail by non-evolutionary theories of biochemistry, cell-physiology, bacteriology, etc. Indeed the theory of natural selection gets its best confirmation in these laboratory settings where it is not actually needed at all. Its confirmation in contexts like paleobiology, ethology, sociobiology, where it is required, is much less precise and detailed.

If the theory of natural selection can only be decisively confirmed in the artificial settings of the laboratory, then it is condemned to perpetual dispute by its opponent. For he wants claims about the evolution of middle-sized organisms mentioned in the book of *Genesis* to be tested with an allowable range of error that would do credit to astronomy; so of course he will remain skeptical. The pro-

[14] See, for instance M. Ruse, *op. cit.,* note 1. The acceptability of laboratory studies for resting the theory of natural selection is defended in A. Rosenberg, "The Evidence for Evolution: Reconsideration," *PSA 1980* (East Lansing, Michigan: Philosophy of Science Association, 1980) 83–93.

ponent of the theory freely employs it, and is at least content that the theory has been tested and never yet disconfirmed. It is no defect in evolutionary theory that it can issue in no more than generic predictions and *post facto* explanations, for this results from the contingent fact that the determinants of fitness are too diverse and complex to permit us to improve the powers of the theory. It would however be a mistake to conclude that no improvements are possible in our explanations and predictions of evolutionary phenomena. But such improvements are forthcoming only by passing beyond evolutionary theory and focusing on the discovery of theories about particular members of the vast and heterogeneous class of determinants of fitness: theories in non-evolutionary, functional biology.[15]

There is a broader methodological moral to be drawn from these conclusions. Consider any scientific theory, T_1, whose explanans-variables can only be identified or measured indirectly by inference from the theory's original explanandum variables. In this case improvements in explanatory and predictive power cannot come through the development of new independent theories about the explanans-variables, for there is no access to these variables independent of their original theory. Improvements in these cases must rely on developing a new theory or theories T_2 that can account for the explanadum variables directly, without appeal to the original theory's explanans-variables. If we are lucky the phenomena will be simple and regular enough to enable us to return to the original theory, applying our new theory to liberate the original explanans-variables from their instrumental dependence on the explanandum-variables. In effect, the explanans variables of T_1 will become explanandum variables of T_2, intervening between T_1 and the explananda of T_1. These T_1-terms may or may not be retained. The terms of electrical theory, voltage, wattage, resistance, provide an example of retained terms. The Mendelian gene, and the mechanical ether provide examples of terms forgone. On the other hand, if the phenomena are so complex that the explanadum terms of T_1 cannot be absorbed in a manageable way into any actual or possible T_2, then those terms and their theories stand in the way of explanatory and predictive improvements. Active coherence with such theories should not be a requirement of theoretical advance, and their

[15] The sense of 'functional' here employed as contrasted with 'evolutionary' is adapted from E. Mayr, *The Development of Biological Thought* (Cambridge, Mass.: Harvard, 1982). My conclusions are strongly at variance with his claims however, for these conclusions suggest that functional biology is the *ultimate source of improvement in* our understanding of evolutionary biology.

terms may be eliminated from subsequent scientific discourse without adverse consequences. In the last section of this paper I shall apply this moral to the vocabulary of intentional psychology.

IV

There are other bodies of theory, whose explanatory terms have the same character as the term 'fitness' does in evolutionary theory. One of these is the body of general statements to which we are committed in our every day explanation of human behavior. We cannot express many of those statements of "folk psychology,"[16] but our willingness to causally explain human behavior by appeal to desires, beliefs, and their cognates reflects the tacit acceptance of such a theory. Many explicit psychological theories share an important property with folk psychology, for they too appeal to intentional states of cognition and sensation in the explanation of human behavior. Their intentionality is evidently a consequence of the apparently representational character of desires and beliefs, and it has been the chief obstacle to a psychological theory that is logically coherent with the study of non-intentional phenomena.

Like the theory of natural selection, folk psychology and its scientific companions have traditionally been taxed with predictive weakness: At best our psychological theories enable us to make generic predictions that cannot be improved or systematically corrected. The cause is that, like the variables of evolutionary theory, those of folk psychology are not specified except by appeal to their effects, and the description of these effects must be so heavily qualified that statements attributing intentional states have no more predictive content than an ascription of a given level of fitness. There are no current prospects of identifying the neural states in which intentional ones are realized, and even when this becomes possible, there is reason to believe they will not be manageable parts of any theory of intentional states. Theorizing about the intentional determinants of behavior will have to continue to rely on behavior itself for all our data. Unfortunately no intentional state by itself determines any particular movement of the body. It is doubtful that it would do so, even if we could hold constant all other relevant intentional states. Our clearest source of data for estimating the intentional states of an agent comes from his verbal behavior. But this class of consequences gives evidence about intentional states that

---

[16] Cf. P. Churchland, "Eliminative Materialism and the Propositional Attitudes," this JOURNAL 78 (1981): 67–90, in which this term is coined, and the associated theory criticized.

requires even more stage setting, correcting, and assumptions about intentional states themselves than the less informative but more reliable movements of the whole body provide. Just as determinations of fitness levels relies on rates of reproduction, qualified and hedged around with *ceteris paribus* clauses about *fitness,* similarly, the determination of what intentional states an agent manifests must appeal to his behavior, but similarly qualified by a wealth of *caveats* about other intentional states of the agent (just because they cannot be held constant). Intentional states or fitness levels are both theoretical states whose occurrence can be established only through appeal to auxiliary hypotheses. And, more important, they can be identified only on the basis of the theories in which they appear because it is not practically possible to develop a nonevolutionary theory of fitness, or a manageable non-intentional theory of intentional states. But while in the one case this suggests that we need to transcend evolutionary theory if we hope to improve our understanding of evolution, in the other case it has often been concluded that to understand psychological states we need a theory that is irreducibly intentional. This claim is mistaken, for it can be shown, by parity of reasoning from the evolutionary case, that the only practicable contexts in which an intentional theory of behavior is decisively confirmed is also one in which the theory, and its intentional variables, are otiose.

Of late, philosophers who argue that a theory of psychological states must be intentional do so by attacking models of psychological processes produced by reflection on artificial intelligence and construction of computer simulations. The general strategy of intentionalist criticisms of these models is to assume that human psychological states are intentional, show that the states of simpler systems are not, and conclude that therefore a latter can shed no light on the former. Arguments of this type rest on a fallacy: that of supposing that a theory which makes no mention of a given phenomenon can tell us nothing useful about that phenomenon. In the context of evolutionary theory this fallacy is revealed by laboratory experiments in natural selection. There it is clear that though the theories that explain what occurs in the petri-dishes or fly-bottles make no mention of fitness, they nevertheless explain the evolution taking place in them. If theories which did not mention fitness could tell us nothing about evolution, that would close off the only avenue open to improving the depth and precision of our theories of evolution. The theories that can do this are precisely the evolutionary ones on whose variables fitness supervenes.

The fallacy in the case of intentional psychology seems to have

been committed in the so-called "Chinese room" *Gedanken*-experiment of John Searle.[17] He defends three conclusions:

> (1) The explanation of how a brain produces intentionality cannot be that it does it by instantiating a computer program. (2) Any mechanism capable of producing intentionality must have causal powers equal to those of the brain. (3) Any attempt to create intentionality artificially could not succeed just by designing programs but would have to duplicate the causal powers of the human brain . . . .
>
> And that is why strong [artificial intelligence] has little to tell us about thinking.[18]

The plausibility of Searle's arguments for this conclusion hinges on the fact that appeals to intentionality are superfluous in explaining computer performance in just the way appeals to fitness are in explaining the course of evolution in a petri-dish. In an artificial intelligence laboratory we may describe and explain the behavior of a chess-playing machine in intentional terms. Doing so will not enable us to predict its middle game play any better than we can predict the middle game play of a human agent. But if the computer is simple enough to have only a small disjunction of internal states that jointly and alternatively realize the intentional states we attribute to the machine, then employing computing theory we can predict its further states, "the moves it makes," to much higher degrees of precision, and we can eventually give these further states an intentional reading as well. Under these circumstances some will say that the computer has psychological states, albeit simpler in physical character than our own. Others, like Searle, will of course reject this claim. But both Searle and the proponent of machine simulation will agree that the intentional description is superfluous to the reliable and precise prediction of what the experimental chess playing computer is doing. But the

[17] "Minds, Brains and Programs," *Behavioral and Brain Science*, 3 (1980): 417–459 cf. also "Analytic Philosophy and Mental Phenomena," *Midwest Studies*, 6 (1981): 405–423.

[18] Searle, "Minds, Brains and Programs," 417. Strong artificial intelligence, which Searle distinguishes from weak artificial intelligence, is the claim that "computers given the right programs can be literally said to . . . have . . . cognitive states." Weak artificial intelligence requires only that computers be useful in testing psychological explanations. The distinction raises the question of whether a justifiable line can be drawn between computer runs which strongly confirm a theory of cognition that animates their programs, and computer runs that instance the cognition itself. I suspect the theoretical distinction is no more viable than that between artificial or laboratory and natural selection. If it were, Searle could accept the argument to follow as showing the worth of weak artificial intelligence, while still denying its bearing on the strong version. cf. Rosenberg, "The Evidence for Evolution: A Reconsideration" note 14 above.

chess-playing computer is to intentional psychology what the bac-
teria-laden petri-dish is to evolutionary theory. Both provide a set-
ting where psychological and evolutionary theories can respectively
be confirmed, and indeed improved upon in precision and detail.
But they do this just because what happens in these settings can be
explained without appeal to these two theories. This fact explains
the plausability of arguments like Searle's.

But the parallel also shows the fallacy in arguments based on the
fact that intentional psychology is otiose in the artificial intelli-
gence laboratory. The *non sequitur* becomes evident when the ap-
propriate evolutionary substitutions are made in the expression of
Searle's conclusions:

> (1) The explanation of how a species acquires a level of fitness cannot
> be that it does so by instantiating the protocol of a laboratory experi-
> ment. (2) Fitness levels of the sort that obtain on earth must be the re-
> sult of as many and as complex forces as have operated on the earth.
> (3) Any attempt to create a given fitness-level artificially would have to
> duplicate all the factors which are causally responsible for its natural
> occurrences. Therefore laboratory experiments have little to tell us
> about evolution.

All three of these claims are unreasonable objections to the rele-
vance of laboratory studies in the confirmation of the theory of
natural selection. If intentionality is like fitness, then both the argu-
ments against laboratory "simulation" of natural selection, and
against computer simulation of intentional phenomena are argu-
ments from ignorance. They trade on the infirmities of the theory of
natural selection and intentional psychology; that is, they trade on
the overwhelming complexity of the factors that underlie the reali-
zation of a given level of fitness and of a given intentional state,
and our ignorance of how these factors aggregate. These arguments
from ignorance suggest that in exceptional circumstances restricted
enough for the complexities to be managed and ignorance to be
removed, the infirm theories do not come into play at all. But this is
a mistake. Although in these cases the infirm theories give away for
explanatory and predictive purposes to special theories of greater
but narrower precision, evolutionary and intentional theories are
still confirmed in these settings. Accordingly, it should not be con-
cluded from arguments like Searle's that the study of programmed
computers "has little to tell us about thinking." These studies have
nothing in general to tell us about intentionality, but from this it
doesn't follow that they have nothing to tell us about thinking. It no
more follows in this case than it follows from the fact that biologi-

cal experiments have nothing in general to tell us about fitness, that they also have nothing to tell us about evolution.

There is a more significant conclusion to draw from this parallel between intentional psychology and the theory of natural selection, a conclusion about the most fruitful course to be taken by a study of thinking, and other intentionally characterized phenomena. We cannot reasonably have higher hopes for an intentional psychology than for a theory of natural selection. Evolutionary theory starts with two advantages: it embodies explicitly stated generalizations to which all biologists accord nomological force; it has already been successfully linked to non-evolutionary theories about objects in its domain, and these give it some limited promise of increasing its predictive content. Intentional psychology embodies no agreed laws or approximations to them, and has not yet been linked to any non-intentional theory about intentional states that will improve its chances of increased predictive content. Without such theoretical companionship, it cannot be expected to break out of the predictive limitations to which all parties now agree it is subject. This means that whether or not intentional theories come ultimately to bear fruit, psychology needs to pursue non-intentional theories of *behavior* hitherto explained intentionally, in particular it must pursue neuroscientific theories. Given the character of intentional terms, progress in psychology is not to be expected through the discovery of neural correlates of intentional states, but by ignoring any such connection and searching for neural correlates of behavior. It is unlikely that such research will eventuate in any improvement of our intentional theories of action, though it is just barely possible. On the other hand, this is the only avenue of inquiry that stands a chance of doing so. Anyone who holds that our intentional psychological theories are inadequate and need to be improved is willy-nilly committed to elaborating thoroughly non-intentional theories of human behavior. If there really are establishable generalizations of intentional psychology, only a non-intentional theory can enable us to identify them. And if there are no such laws of intentional psychology, then our only recourse in the attempt to understand psychological phenomena is the development of non-intentional theories. Either way it appears that the most important and the most promising research strategies in psychology will be neurological ones, and these theories must vigorously forgo the language of intentional psychology.

<div align="right">ALEXANDER ROSENBERG</div>

Syracuse University