

## BETTER THEORIES\*

HOWARD DARMSTADTER†

*University of Massachusetts at Boston*

It is argued that a better theory neither (I) proves better at enabling us to *realize our goals*, nor (II) enables us to make more accurate predictions than a worse theory. (I) fails because it, tacitly, erroneously assumes, in talking of *our goals*, that individual preferences for theories can be aggregated into a social preference ordering; (II) fails because it cannot distinguish between important and unimportant predictions. Neither of these failures can be patched up by appealing to the notion of a true theory. The conclusion is that we as yet possess no adequate understanding of the relationships among theoretical progress, goal realization, and predictive accuracy.

Some theories are better than others. It is generally held that when a theory *A* is better than a theory *B*, one or both of the following are true:

- (I) Generally speaking, a person who uses theory *A* will be more likely to realize his goals than if he were to use theory *B*.
- (II) Predictions based on theory *A* will, in general, be more accurate than will predictions based on theory *B*.

Without the “generally speaking” and “in general” qualifiers, (I) or (II) might be taken as a characterization of the concept of a better theory within a certain philosophical view that might loosely be termed “pragmatist.” With the qualifiers in place, most people would accept (I) and/or (II) as simply stating important facts about theories: As our theories improve, so must our ability both to accurately predict events and to bend nature to our wills. Nonetheless, both (I) and (II) are inadequate even as statements of fact. Both statements conceal difficult conceptual problems, most of the work of camouflage being done by those innocuous looking qualifiers. The bulk of this paper will be devoted to spelling out the problems involved. In the final section, I shall speculate very briefly about possible alternatives to (I) and (II), but in the main, the results arrived at in this paper will be rather negative.

**1. Theories.** By ‘theory’ is not meant just any set of propositions. To be a theory, a set of propositions must contain certain lawlike propositions stating regularities governing a wide range of possible situations. When one uses a theory in a situation covered by the theory’s lawlike propositions, these propositions will—in combination with various propositions stating the particular circumstances of the situation in question, and after certain calculations have been made—suggest<sup>1</sup> various additional propositions about the situation in question. These additional propositions

\* Received October, 1973.

† I would like to thank Howard Cohen, Jonathan Rubinstein, Peter Unger, Lewis Wurgaft, and the referee of *Philosophy of Science* for their comments and criticisms.

<sup>1</sup> The word ‘suggest’ has been used to avoid committing myself to any particular view of the relationship between a theory and its consequences.

often suggest in turn that the employment of various means will be effective in the attainment of certain goals which one may wish to realize in the situation in question. It is natural to suppose that if *A* is a better theory than *B*, then the means suggested by the propositions suggested by theory *A* will prove more often effective than the means suggested by the propositions suggested by theory *B*.

Theories are often joined together to form larger theories. The grander theory will consist of a fairly large number of propositions, various subsets of which form distinct (though not incompatible) theories in their own right. Weaker versions of (I) and (II) might see (I) and (II) as holding of the larger theory, but not necessarily of every component theory. Thus, it might be argued that we constantly endeavor to maximize goal realization or predictive accuracy over the whole sweep of our "total theory," even though this may reduce the predictive accuracy, or prospects for goal realization, in certain subregions of the larger domain. But whether (I) and (II) are understood in a strong or weak sense, they will still prove unacceptable.

**2. Theories and Goals.** The argument against (I) begins with a simple observation: Suppose *A* is a better theory than *B*, and suppose that a person finds that he is more likely to realize his goals by using *A* than by using *B*. This still furnishes no guarantee that a different person, with different goals, would also be more likely to realize his goals by using *A* than by using *B*; for this other person, theory *B* might offer substantial advantages in terms of realizing his goals. Put slightly differently, the mere fact that theory *A* might be more useful than theory *B* in the realization of one set of goals does not rule out *B*'s being more useful than *A* in the realization of some other set of goals.

The point I am making is rather simple, but its understanding might be aided by a little story which will also prove helpful at a later juncture. Suppose there are two travelers, taking different routes through an unfamiliar land, each wishing to stop each evening at a good motel. Each traveler has some old guide books and road maps which provide him with information about roads and motels, but each is aware that his information may be seriously out of date—new roads may have been built and old ones torn up, and new motels may have been built and old ones changed or destroyed. Each traveler constantly revises his maps and guide books during his travels, and each picks up some sketchy beliefs about why and where highway departments build and destroy roads in the way they do, and why and where motel magnates build and destroy motels in the way they do. These beliefs allow each of them to estimate just when their revised maps and guide books are likely to prove inaccurate. Each traveler thus develops his own theory of roads and motels, and each uses his theory to act so as to achieve his goals.

Now suppose our travelers have differing preferences for motels. Each comes to possess a theory which proves moderately successful in getting him to a motel of the preferred type. But if each used the other's theory, neither would often reach a motel which he deemed acceptable. Nor need there be any other goals possessed by the travelers whose realization would be differently affected by the choice of one theory over another; for example, it might be the case that if either traveler used the other's theory, he would often take different routes from point to point, but would

lose on the average about the same time in traffic jams, detours, etc. In such a case, the only difference between the theories in point of practical consequences for the travelers would lie in the types of motel in which they find themselves lodging. So it may be that one theory is as good as another given the differing preferences of its users. Yet each theory could suggest quite different propositions about roads and motels.

**3. The Aggregation Problem.** Even though  $A$  is a better theory than  $B$ , it might be that for some individuals,  $B$  is more useful for the realization of their goals than  $A$ . Still, this does not by itself show (I) to be inadequate. While  $B$  may be a more useful theory than  $A$  for some people, it might be argued that  $A$  will be more useful for more people or for the realization of more goals than will  $B$ ; this, one might add, is the meaning of the “generally speaking” qualifier in the statement of (I).

People’s goals differ, and since one theory may be more useful in the realization of some goals while another theory may be more useful in the realization of other goals, we may expect people to differ in their preferences among theories. For each person, with his individual set of goals, there will be a ranking of theories in order of their usefulness in the realization of that person’s goals. What is needed is some way of *aggregating* these individual preference orderings into some sort of social preference ordering. If  $A$  is a better theory than  $B$ , we would expect it to stand higher in the social preference ordering, regardless of where it stands relative to  $B$  in any individual preference ordering.

The above attempt to save (I) thus depends on our being able to aggregate individual preference orderings (in this case, preferences for theories) into a social preference ordering. But any such attempt would run afoul of Kenneth Arrow’s famous “Impossibility Theorem” for social welfare functions. Briefly stated, Arrow proved that there is no way, in general, of aggregating individual preference orderings into a social preference ordering without violating certain common-sense restrictions designed to make such aggregating procedures reasonable. Put slightly differently, Arrow proved that there is no way of simultaneously satisfying all of four minimal conditions on the aggregation of individual preference orderings. The conditions are:

(1) There must be a social preference ordering for every logically possible configuration of individual preference orderings.

(2) If everyone in the society prefers alternative  $x$  to alternative  $y$ , then society must prefer  $x$  to  $y$ , i.e.  $x$  must be ranked higher than  $y$  in the social preference ordering.

(3) The social preference ordering among a set of alternatives must depend solely on the individual preference orderings among *those* alternatives. Thus, if  $x$  is socially preferred to  $y$ , and if a new alternative  $z$  is introduced, then as long as no one changes his individual preference ordering between  $x$  and  $y$ , then  $x$  should still be socially preferred to  $y$ , regardless of where  $z$  is ranked. If, for example, society prefers vanilla to chocolate in a two-way race, it should still prefer vanilla to chocolate in a three-way race, regardless of where it might rank strawberry.

(4) There is no individual such that whenever he prefers  $x$  to  $y$ , society must prefer  $x$  to  $y$  irrespective of the individual preference orderings of the other members of the society.<sup>2</sup>

**4. True Theories.** Arrow's theorem does not present an insuperable barrier to the aggregation of individual preference orderings. If we are willing to introduce certain special assumptions about the nature of each individual's actual preference ordering, i.e. if we are willing to weaken condition (1) above, then we may be able to produce a method of aggregating individual preference orderings which, while it will not prove acceptable when applied to certain logically possible sets of preference orderings, will prove acceptable when applied to the limited types of preference orderings which we may feel are those actually possessed, or ever likely to be possessed, by human beings as presently constituted. Unfortunately, no one has been able to produce a set of assumptions about individual preference orderings which seem likely to be true and which, at the same time, yield a social preference function which can make certain minimal distinctions among society's preferences for differing alternatives; for example, many of the social preference orderings arrived at with the use of special assumptions make all Pareto-optimal outcomes equally desirable.<sup>3</sup> It may be, however, that in the case of individual preference orderings of theories, certain special assumptions may prove acceptable which would not be acceptable when considering individual preference orderings among other sorts of entities. I do not wish to consider here all the special assumptions that a philosopher might wish to make in order to aggregate individual preference orderings among theories. Instead, I will limit myself to one quite special assumption which people have usually been quite ready to grant, namely that there are certain theories—the *true* theories—which could be judged to be socially preferred to all competing theories simply because they would be preferred by every individual. Such theories would thus constitute a special case which would avoid the general problems raised by Arrow's theorem. Statement (I) would still be inadequate when the better theory referred to was not a true theory, but it would be adequate where  $A$  was not merely a better theory but the best, i.e. the true, theory. It would follow that anyone using a true theory will be more likely to realize his goals than he will if he uses any non-true theory covering the same situations. Thus, in the story of the two travelers, each of our travelers would have been better off (it is argued) if they had both used a (or the) true theory, which theory would presumably have told them where all the roads were, which motels had color TV, etc. Using such a true theory, each of our travelers would have been as likely to reach a motel he deemed acceptable as he would have been using any nontrue theory.

Admitting for the moment that there are true theories, does it follow that using a (or the) true theory will better enable one to realize one's goals than will using a false theory? I think not. Consider our two travelers: Each employs a primitive

<sup>2</sup> Accounts of Arrow's work, in ascending order of difficulty, are provided in [1], [2], pp. 327–345, and [4], pp. 1–46. This particular set of conditions is taken from [4], pp. 37–38, 41–42.

<sup>3</sup> See [4], pp. 41–200, and especially, pp. 67–77, for a full treatment of these issues. A more limited account is given in [2], pp. 345–370.

(and to some degree false) theory of map-making and map-conjecturing. Far away, in the lavishly staffed and equipped offices of the atlas maker, a true theory of map-making and map-conjecturing is used. The *conclusions* which the atlas maker derives from his true theory would be of great value to our travelers, but the mere possession of the theory might be of no use to our travelers, whose devices, supplies, and communications are relatively meager. If using a true theory requires special equipment, or a large staff, or a great deal of time, all of which are unavailable to our travelers, then their possession of that theory may do them no good at all. Thus, for some people, a simple false theory might be better than a cumbersome true one.

It may seem natural to many readers to identify "better theory" with "truer theory." Such an identification seems somewhat dubious as an attempt at explication—generally speaking, the notion of a better theory has usually seemed simpler than that of a truer theory. But however unilluminating such an identification might be, will it solve our problem? In particular, will (I) hold where *A* is a truer theory than *B*? It is not at all clear that it will. As was demonstrated above, a "somewhat false" theory may prove more useful than a truer theory if the truer theory is more cumbersome to use. One might admit this, but argue that (I) merely requires that "*generally speaking*, a person who uses *A* (the truer theory) will be more likely to realize his goals than if he were to use theory *B*." (I) might hold when *A* is a truer theory than *B* if we can assume that, in general, truer theories are no more complex than less true theories. I do not see how one might conclusively demonstrate this without knowing (what we do not know) how to rank theories as to degree of truth. There does, however, seem to be some evidence against such an assumption: As I shall argue in the following section, it seems that the truest theories of the world may be among the least useful in the realization of our goals.

**5. Macro and Micro.** In presenting the example of the two travelers, I was not intending to suggest a mere logical possibility to the effect that under certain circumstances it might be better to possess a false theory than a true one. It seems that certain fundamental beliefs about the universe, held by most educated men, have the consequence that our normal situation in the world is virtually always closer to that of the two travelers than to that of the atlas maker.

Most of us tend to believe that the basic laws governing the universe apply to very small things. The macro-objects of our everyday experience are composed of a vast number of these micro-objects, and the behavior of the macro-objects is determined by the interactions of the many micro-objects which compose the macro-objects. Any of the events of our experience is the result of a vast number of very small events which "add up" to the event we witness.

Micro-events are governed by laws, and a statement of those laws provides a true theory of the way the world works. Macro-events can in turn (at least in theory) be predicted by using the laws governing micro-objects and micro-events, adding a great deal of information concerning various properties of particular micro-objects at the relevant times, and doing some elaborate calculations. But the number of micro-objects involved in even the smallest events of our experience is so vast that giving a complete list of the doings of all the micro-objects involved and per-

forming the relevant calculations becomes a practical impossibility. There are, of course, lawlike (and practically manageable) regularities which apply directly to macro-objects, and we constantly use such regularities in the governance of our daily lives. But we believe that predictions made on the basis of regularities applying directly to macro-objects will be less accurate than would predictions based on the "real" laws governing micro-objects. Indeed, if the laws governing micro-objects did not allow (in theory, and occasionally in practice) more accurate predictions of macro-events than a manageable system of laws applying directly to macro-objects, then we would never have assumed that macro-objects and events are produced by a vast number of micro-objects and events.

If we believe that a true theory is one which enables us to make (in theory, if not in practice) predictions more accurate than can be made with any nontrue theory, and if we believe that the most accurate predictions which are (theoretically) possible involve laws governing micro-objects, then true theories will be theories about micro-objects and micro-events, and such theories will be virtually of no use to us in most situations. Nor can the development of computation procedures and computing mechanisms significantly widen the range of such theories' applicability. For the best predictions will involve information about all the micro-objects in the universe (anything less raises the not-altogether-theoretical possibility of error), and such predictions must always remain a merely theoretical possibility.

**6. Predictive Accuracy.** Does (II) hold? That is, if  $A$  is a better theory than  $B$ , must predictions made on the basis of  $A$  be generally more accurate than predictions made on the basis of  $B$ ? No, for reasons that should now be obvious. Any theory about macro-objects will be inaccurate to some degree, and different theories may differ as to their degree of accuracy in various areas. A rational theory constructor will try to have the greatest (practically attainable) predictive accuracy *in those areas where it is most important* to him to have accurate predictions, and will require less predictive accuracy in those areas where he can afford to do without accurate predictions. If two rational theory constructors have different goals, then they may not agree on those areas in which it is most important to have accurate predictions, and those areas in which substantial inaccuracies can be tolerated. The test of a theory for an individual is therefore not how much predictive accuracy it has, but how much predictive accuracy it has in those areas where the individual in question deems it important to have accurate predictions. Thus, if theories  $A$  and  $B$  differ in respect of which areas predictive accuracies and inaccuracies fall, then different rational theory constructors may rationally choose different and incompatible theories. Some way may of course be found to combine the two theories into a new theory which is as accurate in any area as either of the original theories. But there seems no *a priori* reason to believe that two individually manageable theories can always be combined into a manageable theory.

At the present time we possess theories which enable us to make certain predictions concerning various ranges of phenomena in which we have an interest. In particular, we can make fairly accurate predictions concerning the performance of various complexly constructed artifacts—television sets, digital computers, etc.

Practical men regard these successful predictions as showing the superiority of the current theories over any alternate theories which would fail to make such predictions, or fail to make them with the same degree of accuracy. But it remains possible that the alternate theories that practical men discard might be preferred by people who were less interested in television sets and digital computers, and more interested in ranges of phenomena concerning which the current theory offered no predictions or less accurate predictions than the alternate theory. Virtually every theory has some evidence for it, and virtually no theory can be conclusively disproved by the failure of any set of predictions.<sup>4</sup> Nor can the superiority of one theory over another be demonstrated by appeal to a larger quantity of evidence, for no way is known to quantify bodies of evidence. Moreover, it seems reasonable to expect rational theory constructors to prefer a theory with a fairly small amount of evidence if that evidence is in the form of accurate predictions in an area where it is extremely important to have accurate predictions, even though the theory in question may provide inferior predictions in a fairly large number of areas deemed to be of little importance. There is nothing wrong with preferring current physical theory in part because of its predictive accuracy with respect to the construction of television sets and digital computers, but we should realize that such a preference is based in part on the fairly large positive valuation we accord to the goal of having reliable televisions and computers.

**7. Conclusion.** Despite Arrow's theorem, I still prefer democracy to dictatorship, and prefer certain democratic choice procedures over others. Arrow's theorem shows me that I cannot justify these preferences by any appeal to a "general will" which can be amalgamated out of individual preferences. But my preference for democracy over dictatorship may still depend on the very different ways in which the preferences of individuals enter into democratic and dictatorial choice procedures. Obviously, the problem of explaining the superiority of one system over another is a much more complicated one than most people imagined before Arrow. Arrow's theorem does not, however, make the justification of such preferences impossible, nor does it even force us to give up all hope of a justification that is based on the relation of social choice procedures to individual preferences; there need not exist a general will in order for us to say that democracy is preferable to dictatorship because the preferences of more individuals have an actual relevance to the final social choice.<sup>5</sup>

Similarly, the rejection of (I) and (II) should not force us to deny what seems an obvious truth—that the superiority of one theory over another must have *some* relation to predictive accuracy and helpfulness in realizing goals. But the relation between the accuracy of our predictions, the ability to realize our goals, and the superiority we feel certain theories to have over others will be much more complex than was suggested by (I) and (II). For example, it may turn out that the structure of various social institutions—science, government, the media, etc.—may give disproportionate importance to certain goals of certain people, and that it will be possible to specify a relation between the better theory and the goals of this small

<sup>4</sup> The second conjunct relies for argument on W. V. Quine's well-known essay, [3].

<sup>5</sup> Theorem 5\*1 of [4], p. 72, is particularly interesting in this connection.

set of people. But whatever form the relation might take, specifying it is going to require a good deal of work, and it may be that much of this work will fall within the province of disciplines other than philosophy. But neither hard work, nor the prospect of having to seek the answer in other disciplines should cause us to ignore the problem.

## REFERENCES

- [1] Arrow, K. J. "Values and Collective Decision-Making." In *Philosophy, Politics and Society*. Edited by P. Laslett and W. G. Runciman. Oxford: Basil Blackwell, 1969. Pages 215-232.
- [2] Luce, R. D. and Raiffa, H. *Games and Decisions*. New York: Wiley, 1957.
- [3] Quine, W. V. "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, Massachusetts: Harvard University Press, 1961.
- [4] Sen, A. K. *Collective Choice and Social Welfare*. San Francisco: Holden-Day, 1970.