KYKLOS, Vol. 53 – 2000 – Fasc. 4, 587–592

How to Make Scientists Agree: An Evolutionary Betting Mechanism

Tilman Slembeck*

I. INTRODUCTION

One of the basic problems in giving economic advice is to find sufficiently robust empirical evidence for making policy recommendations. In many fields there exist conflicting theories and empirical evidence is often mixed. For instance, theories on economic growth, innovation, and technology or on the regulation of competition (anti trust laws) have devised results that are not always robust, difficult to condense into simple recommendations, and sometimes contradictory. The source of conflicting results in many areas are to be found in varying sets of assumptions in theory and different standards and methods in empirics.

In order to base policies on "sound economic knowledge" (or indeed sometimes only to legitimize already taken decisions and to support existing preconceptions and beliefs) policy-makers employ experts. The task for expert advisors or advisory boards is to base their recommendations or proposals on some „general findings" or „stylized view of the facts" (Kaldor, 1968) in order to legitimize their advice. This task, however, often turns out to be difficult due to the mentioned problem of conflicting theoretical and empirical results in many (though not all) fields.

Although there appears to exist some basic consensus with regard to some issues (see Alston et al., 1992; Frey et al., 1984; Kearl et al., 1979), this consensus is not overwhelmingly broad, is to be found only at a rather general level of abstraction, and tends to fall apart when it comes to policy recommendations. Hence, the joke that by asking two economists for advice you get (at least) three

* Department of Economics, University of St.Gallen, Varnbuelstrasse 19, 9000 St.Gallen, Switzerland, E-mail: tilman.slembeck@unisg.ch. I am grateful to Al Roth for predicting that "it is hard to upset a well established system for recognizing and rewarding scientific work with a new, alternative system."

different answers. This situation leads to the more general question of how it may be possible for (economic) scientists to agree on some scientific result.

Since there do not exist any "objective" ways to decide on the "truth" of scientific results, social (and other) sciences have resorted to some method of "temporary consensus." This has lead to various schools of thought involving scholars that agree upon certain views, assumptions and methods. While there may be agreement within the school, there appears to be much disagreement between schools. Although the scientific competition between schools may have produced "scientific progress" in terms of the number of theories and sometimes in terms of robustness and the range of applicability of results, there remain to exist substantial and basic differences in focus, assumptions and methods – and hence in "results".

In the following I will try to outline some basic mechanism that may help to bridge some of the disagreement between schools, but especially within schools insofar as scholars are able to agree on empirical methods to test conflicting theories or hypotheses. The aim is to enlarge the area of consensus on „general findings" or „stylized facts" with regard to policy-making. The prerequisite, however, is that scientists, possibly even from different schools, are willing to engage in direct empirical competition, make available empirically testable theories or hypotheses, and are able agree on empirical methods to test them. The incentives involved with the mechanism once it is implemented, however, may encourage scientists to meet these requirements.

II. POPOSED MECHANISM

The basic idea is to *involve* two scientists holding conflicting or opposing views, theories or hypotheses on a given topic or research question *in a scientific competition* or "*bet*". That is, two scientists (say, theorists) bet on their prediction or hypothesis to hold or be supported by the data with regard to a certain phenomenon or research question. A third (i.e. empirical) scientist is employed to do the empirical work. The challenge for this third scientist is to propose a method that allows to reject the hypotheses, and that both bidders can agree upon. After such agreement is reached, the empirical part is executed by the third scientist alone without any interference of the two competitors, and it is announced who won the bet. The entire procedure and results are then published as a joint work of all three scientists involved.

The basic incentive for scientists to engage in such betting may be provided by a public scoring or ranking of scientists. That is, the scientist winning a bet

receives a certain number of "points" on the *academic ranking list*. Such a list would *complement* the existing methods of ranking scientists (or academic departments) based on publications or citations. Hence, the academic standing or reputation of a scholar would *also* be gauged by his or her ranking on that list. Since the proposed mechanism submits scientists to rather strong conditions and procedures, there may be a high reputation of such a list, and hence strong incentives for being on the list. It also provides incentives for theorists and empiricists to work together (see also Hofstee 1984).

In order to ensure the quality of the work publication is required for authors to receive any "points". The number of "points" one receives for a won bet may also depend on the ranking of journal where the paper is finally published so that the publication in a highly ranked journal yields comparably more "points". That is, some "basic points" for winning are multiplied with a factor associated with each academic journal.

The design of the incentive scheme in terms of "points" (P) may also involve appropriate incentives for empiricists to "run the bet" and for losers. For instance, winning a bet may give 2P and running it yields 1.5P, while losing it still would give 1P. Giving "points" to losers is to reward them for joining the bet, and to provide incentives for participation in such betting (which, however, must be published to yield "points"). Also, winning a bet against a hypothesis that has previously won other bets, or against opponents of high ranking may yield "extra points" (similar to the ATP ranking in tennis).

In cases where both hypotheses under investigation are empirically rejected no one wins so that both losers receive the same default number of "points" for joining the bet, since the result may still be of scientific value (to be judged by journal referees). In case that no hypothesis can be rejected the empirical method is likely to be flawed so that no "points" are awarded, and the work is probably difficult to publish. This leads us to the case where one (or none) hypothesis won the bet, but there are doubts about the empirical method. In this case a new bet among empiricists may be run.

III. DISCUSSION

The proposed mechanism is based on a betting-market with external incentives. These incentives call on the scientists' strive for reputation, and the implied effects on academic status, income, and access to research funds and facilities. It is a pairwise procedure, although it may sometimes be possible to involve more than two scientists in betting. Ideally, winners are challenged by new contestants who aim to increase their reputation. Such multi-stage betting is *evolutionary* in

nature in that it does not require some authority to guide or direct the process (aside from the incentive scheme). The result of this process is open in the sense that the path that consecutive betting takes is not predetermined. But the process is also to some extent path-dependent in that later bets are based on the results of earlier bets the outcomes of which limits the space of hypotheses available in a given research area. However, since new results often tend to create new questions and puzzles, the process is unlikely to "converge", but to produce new results based on temporary agreement.

Agreement is important in several respects, because in order to run the bet the scientists must agree upon several things up-front. First, they have to agree on the issue or research question that has to be answered. The question has also to be defined sufficiently precise or specific in order to be made operational for empirical testing. Second, the scientists must make available two (or possibly more) empirically testable hypotheses that allow for sufficiently different predictions. Obviously, they have to agree that each others' hypothesis is worth testing. Third, the scientists have to agree upon the empirical method that will be employed to test both hypotheses adequately. This may, for instance, involve agreement on what data set or what experimental design to use. At this stage empiricists may help by proposing new and innovative methods. Fourth, scientists must agree on a third scientist who executes the empirical research (and he or she must agree on doing so). – The incentives for reaching such agreements may be strong once an accepted scoring list is in place. Given these ex ante agreements, it is difficult for scientists to talk their way out of the evidence ex post when the results do not fit with their predictions or hypotheses.

The method also opens new opportunities for *specialization*. In cases where scholars of different schools "don't talk to each other" other (young) scientists may specialize in confronting theories in the way described. Similar arguments may hold when theories are too general or abstract for empirical testing. Hence, there may be additional incentives to deduce testable hypotheses from "pure" or normative theories once the mechanism is in place. Empirical research institutions may specialize in "running the bets", and new journals (or sections of existing journals) may specialize in publishing them.

The proposed mechanism appears to have straightforward desirable features in that it provides incentives to use scarce research resources more efficiently with regard to an increase in number (and possibly a reduction in variance) of empirically tested theories. Such redirection of research resources toward the confrontation of conflicting theories and empirical testing may indeed be desirable only if we understand economics as a theoretical *and* empirical discipline, and if we believe in the value of empirical work as a basis for scientific progress and policy advice.

Although many economists may subscribe to this view (as I assume), the proposed mechanism may still be difficult to *implement*. First, there do already exist established mechanisms for reputation-building. The mechanism, however, is not supposed to replace them but to complement them, and to add incentives in the direction just described. Second, scientists may avoid being confronted with others directly (although conventional scientific moral seems to suggest otherwise). The most promising way for implementation would probably be if the mechanism was adopted by some major economic association (such as the AEA or EEA), starting with a bet between two eminent scholars on an issue of basic importance and general interest so that the procedure would be in the center of attention of the scientific community. The association may then start to publish the personal rankings and routinely award prices (e.g., for certain periods or for "life-time achievement").

If the proposed mechanism (or variants to be discussed) cannot be implemented even though its potential benefits once it is in place are generally acknowledged by the scientific community, economists may still learn something about themselves and about economics. First, implementation of "beneficial" (say, efficiency-enhancing) mechanisms can be difficult or impossible even if many agree on its benefits compared to the status quo. Hence, it is often not enough to design optimal or efficient mechanisms and leave the problems of implementation to others. Second, most people tend to avoid direct (or in fact any) competition if they can. Business leaders praise the benefits of free markets and competition in their public statements, but have devised large arsenals of tools to reduce it in the markets they deliver. Economist have stressed the positive welfare effects of competition in the economic realm, but may avoid direct confrontation when it comes to their own business.

IV. CONCLUSION

The aim of the proposed mechanism is at finding agreement on empirically robust results that can be used for economic policy advice. The mechanism cannot dispense advisors from deducing policy implications from the „general findings" or „stylized facts", but it may provide them with a somewhat sounder basis for doing so. This is the only way to proceed if we believe in the value of a combination of theory and empirical work as a basis of economic policy advice – as compared to "pure", or normative, or no theory.

The mechanism does by no means guarantee that policy-makers will adopt economists' proposals more thoroughly. Policy-makers are still likely to stick to their own views, beliefs and preconceptions. They will still tend to use advisors

to support their own interests, and to ignore advice that runs counter to their normative and positive beliefs. However, economists may have a somewhat stronger argument once such a mechanism is able to increase the degree of consensus within the discipline on certain issues or fields of research.

In the light of the mentioned difficulties economists may even be tempted to refrain from giving policy advice altogether. However, this would mean leaving policy-makers to their own, which may not be a too attractive alternative. And, after all, there appears to be a demand for policy advice that provides economists with many well-paid jobs.

REFERENCES

Alston, R.M., Kearl, J.R., and Vaughan, M.B. (1992). Is There a Consensus among Economists in the 1990's?, *American Economic Review*. 82: 203–209.

Frey, B.S. et al. (1984). Consensus and Dissension Among Economists: An Empirical Inquiry, *American Economic Review*. 74: 986–994.

Hofstee, W.K.B. (1984). Methodological Decision Rules as Research Policies, *Acta Psychologica*, 56: 93–109.

Kaldor, N. (1968). Capital Accumulation and Economic Growth, in: Lutz, F.A.and Hague, D.C. (eds.). *The Theory of Capital - Proceedings of a Conference Held by the International Economic Association* (1958), London: Macmillan: 177–222.

Kearl, J.R. et al. (1979). What Economists Think – A Confusion of Economists?, *American Economic Association Papers and Proceedings*, 69: 28–37.