

On the diversity principle and local falsifiability

Uriel Feige

October 22, 2012

1 Introduction

This manuscript concerns the methodology of evaluating one particular aspect of TCS (theoretical computer science) papers in general, and AGT (algorithmic game theory papers) in particular. This aspect refers to the potential practical applications of these papers. There are multiple other aspects that contribute to the evaluation of papers, such as their mathematical content, aesthetic appeal, relations with previous work, and quality of writing. However, these and other aspects, some of which are often more important than the aspect discussed in this manuscript, will not be addressed in this manuscript.

We shall argue that a certain principle, that we call here the diversity principle, makes it difficult to judge the possible practical applicability of TCS papers. We shall then present an evaluation criterion that we shall call *local falsifiability* that helps cope with this difficulty.

1.1 Related work

The current manuscript is mostly motivated by models and assumptions used in AGT papers. Previously, Moni Naor [5] initiated a discussion of falsifiability issues for models and assumptions used in cryptography. The nature of assumptions in AGT papers is different from that in cryptography papers, and hence, despite some similarities, the concerns discussed in the current manuscript are different from those in [5]. In non-TCS communities, there have been discussions of issues related to those discussed here. See [1] for example.

2 The main thesis

2.1 Theoretical versus practical models

In our discussion we shall use the terms *model* and *assumption* interchangeably. For example, we treat statements like *utility functions are modeled as monotone functions* and *utility functions are assumed to be monotone* as equivalent.

We draw a distinction between what may be called a *theoretical model* and a *practical model*. A theoretical model merely has to be well defined. It needs no external motivation (though it may have external motivation). A practical model needs to also capture the essence of some phenomenon from the physical world, either from the exact sciences, or as is more common in AGT literature, from the social sciences (a social phenomenon, an economic phenomenon, a psychological phenomenon). Sometimes, a model can serve both as a theoretical model and as a practical model.

Consider for example the model of rational agents in the sense of von Neumann and Morgenstern [6]. This is an inspiring theoretical model that lead to beautiful theory. In addition to its intrinsic mathematical value, this is a model that one might sometimes want to use in practice. But is this model really practical? This is a question that concerns not only the mathematical nature of the model but also the nature of the real world, and hence is difficult to evaluate using mathematical tools alone.

2.2 Falsifiability

How does one evaluate a practical model? Are there any objective guidelines for this? Here, let us discuss a principle advocated by the Philosopher of Science, Karl Popper. This principle says that in order to be regarded as scientific, a theory (or model) must be *falsifiable*, at least in principle. That is, there should be tests available (e.g., experiments) such that if the theory contradicts the outcome of the tests, one is willing to abandon the theory (or at least acknowledge that the theory needs to be modified). Can we apply the falsifiability principle to practical models in AGT papers?

If an AGT paper introduces a practical model for one particular application, the falsifiability principle would in general apply. It is likely that one could design experiments that test whether the conclusions derived from the model hold in practice, and if they do not hold, the model has been falsified. For example, there were experiments that showed situations in which people do not behave in a way predicted by the above mentioned rational agent model [3, 4], and hence in these situations the rational agent model needs to be abandoned (or modified).

2.3 The diversity principle

AGT papers have become more subtle when putting forward a model, and rather than claiming a particular well defined application, they introduce a theoretical model and say that it *may have* practical applications. Indeed, one may argue that there is no one practical model that fits all situations. Many practical models, possibly conflicting with each other, may coexist, and in any particular application one should choose the practical model that suits it best. For example, it is sometimes convenient to model light as a wave, and sometimes convenient to model light as a particle. Even models that appear not to be realistic may eventually turn out to have practical value in some situations. We call this argument *the diversity principle*.

The diversity principle has some truth in it, and hence it became part of the culture of AGT papers. However, it also has some undesirable aspects. The phrase “may have practical applications” empties the content from the falsifiability principle – no matter in how many situations one demonstrates that the practical model does not apply, the practical model might still apply in some new situation. The diversity principle does not distinguish between good and bad practical models, and hence instead of serving as a guideline for choosing a useful practical model, it might serve as an excuse for accepting useless practical models.

2.4 Local falsifiability

We offer *local falsifiability* as a guideline of how to deal with the diversity principle. Suppose that based on a proposed practical model, one designs a mechanism (in the sense of mechanism design, see [7]) for solving some underlying task (e.g., say for performing a combinatorial auction). The implicit claim of the diversity principle is that there will be some practical situations in which the mechanisms applies. One policy of taking advantage of this implicit claim is to apply the proposed mechanism in every opportunity (e.g., whenever one conducts a combinatorial auction), in the blind faith that the underlying model happens to hold. This may be a good policy if the mechanism offers significant advantages when the model does hold, but does not suffer significant negative consequences when the model does not hold. However, it is often the case that mechanisms are harmful if the underlying practical model fails to hold, and then one would prefer not to employ the mechanism at all unless the underlying assumptions that led to its design are known to hold. This raises the question of whether one can test in (at least some) given practical situations whether the assumptions of the model actually hold. Following the falsifiability principle, the local falsifiability principle would stress testing that a model does not hold, rather than that it does hold. This leads to the following informal definition.

Definition 1 *A practical model is locally falsifiable in a given practical scenario if there is test that can be implemented in the given scenario that could potentially refute the assumptions of the model.*

Rather than trying to formalize concepts such as *practical scenario*, a *test*, or *potentially refute*, let us explain the intended meaning behind our definition using a few examples.

1. Assumption: agents do not communicate with each other. This assumption can in principle be refuted, for example, by monitoring all communications of the agents and observing that some of them do communicate with each other. Hence mechanisms that make this assumption may be practical in certain situations.
2. Assumption: when asked to output a random bit, an agent produces an unbiased random bit independent of all previous bits output by the agent. This assumption may in principle be refuted (or at least significantly discredited), for example, by designing a statistical test before the agent starts outputting bits, and observing that the test correctly predicts the next bit in the sequence for well over half the bits. However, in practice refuting this assumption might be very difficult (which is why pseudo-random sequences are so useful). If a mechanism heavily relies on this assumption (for example, this assumption ensures that the sequence of bits output by the agent does not serve as a subliminal communication channel with another agent), then it would be dangerous to use such a mechanism in practice – it is too easy to violate the assumption without the mechanism being aware of the violation.
3. Assumption: an agent has a utility function in the sense of von Neumann and Morgenstern [6]. This assumption is known to be equivalent to certain consistency conditions among lotteries. By offering the agent various lotteries and observing inconsistencies among the agent's choices of lotteries, one can in principle refute this assumption.

However, the mere testing of this assumption might actually change the utility function of the agent. For example, if the agent has nonlinear utility for money, then by winning some money in a lottery, the marginal utility function for money that results from this win need not be identical to the original utility function. (See [2] for additional aspects related to this issue.) In some cases, this might make the test impractical.

The premise of our approach is that the scope of the diversity principle is limited to scenarios in which there is a local falsifiability procedure.

2.5 Consistency between mechanisms and local falsifiability procedures

It is desirable that in a paper that appeals to the diversity principle, the authors describe potential local falsifiability procedures. This may allow one to judge whether the local falsifiability criterion was demonstrated to hold. This judgement might be quite subtle. Let us illustrate this by a hypothetical example. Suppose one designs a mechanism M_D based on the assumption that utility functions of agents are convex. Suppose moreover that the main appeal of the mechanism M_D is that it is deterministic rather than randomized, and in order to achieve determinism the designer was willing to make M_D significantly more complicated than an alternative randomized mechanism M_R . The designer argues that being deterministic is of utmost importance because one cannot trust randomness in practice. Now, to apply M_D (or M_R) in a real world situation, local falsifiability would require the designer to also supply a local falsifiability procedure for the assumption that the utility functions are convex. Suppose that similar to item 3 in Section 2.4, the suggested local falsifiability procedure is based on lotteries, namely, is randomized. Then this would contradict the main motivation for designing M_D . This brings us to another important aspect of the local falsifiability criterion: the advantages of a new proposed mechanism over other alternative mechanisms should be demonstrated in an environment in which there is a local falsifiability procedure.

3 Discussion

One may offer the following classification of AGT papers in terms of their claims for practical relevance.

1. **Basic theory.** These are papers that do not make any claims regarding practical applications.
2. **Anecdotal motivation.** These are papers that claim to be motivated or inspired by some practical problem, but do not claim that the results proved for the theoretical model translate back to the motivating practical problem.
3. **Practical application.** These are papers that claim to be directly relevant to a specific practical application.
4. **Diversity principle.** These are papers that claim practical relevance, but without committing to any specific practical application.

Our manuscript mostly relates to the evaluation of the fourth class of papers. More specifically, it relates to evaluating only one aspect of these papers (the claim of practical relevance) and ignores other aspects that might be more important (such as mathematical content). What we attempt to do is to establish criteria for evaluating whether the authors managed to substantiate the claim of practical relevance (and hence deserve credit for it), or whether it should be regarded as a superficial unsubstantiated claim. Our proposed methodology suggests that the reviewer first determine whether the assumptions have a local falsifiability procedure. Ideally, such a procedure will be supplied by the authors (this is certainly not common practice in current AGT papers, but hopefully will become so as the field develops further). In some other cases, it might not be difficult for the reviewer to come up with plausible local falsifiability procedures. However, if neither case holds, the conclusion of the reviewer should be that the authors failed to substantiate their claim of practical relevance. Suppose now that the assumptions do have a local falsifiability procedure. The next step is to determine whether given an environment that supports these local falsifiability procedures, would the results of the paper still be relevant (or would one want to do something else in such an environment)? If indeed the results remain relevant, the claim for potential practical relevance has been demonstrated to be plausible. (This does not mean that the results of the AGT paper are actually relevant to practice, but only that it is plausible that they are.)

It would have been instructive at this point to pick a few existing AGT papers of the fourth class above and attempt to evaluate them based on the local falsifiability criteria. However, we refrain from doing this here, and leave this task to the interested reader.

References

- [1] Samuel B. Bacharach. Organizational Theories: Some Criteria for Evaluation. *The Academy of Management Review*, Vol. 14, No. 4 (Oct., 1989), pp. 496–515.
- [2] Uriel Feige, Moshe Tennenholtz. Responsive Lotteries. *SAGT 2010*: 150–161.
- [3] Daniel Kahneman, Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, XLVII (1979), 263–291.
- [4] Andreu Mas-Colell, Michael Whinston, Jerry Green. *Microeconomic Theory*. Oxford University Press 1995.
- [5] Moni Naor. On Cryptographic Assumptions and Challenges. *CRYPTO 2003*: 96–109.
- [6] John von Neumann, Oscar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, Princeton, third edition 1954.
- [7] Noam Nisan, Amir Ronen. Algorithmic Mechanism Design. *Games and Economic Behavior* 35(1-2): 166–196 (2001).