Advice, Critique, and a Technical Topic

Oded Goldreich

June 21, 2017

After thought: A few such comments were added after presenting the talk.

1 title slide

1. Non-Technical: Advice and Critique.
2. Technical: On Doubly-Efficient Interactive Proof systems

Having turned sixty in February, I now belong to the elders. So I may be pardoned for promoting a good old vision, and arguing that it is still relevant, and maybe even more relevant than ever before. But apologies come first...

2 Apologies

1. For lying (to people who asked me of my attending STOC'17).

   Cf., finance ministers lying re currency devaluation before the monetarization coup.

2. For not declining the award (as some might have hoped...).

   I object to creating them and to being preoccupied with them, since they serve no common good and create artificial hierarchy, but given that they exist I see little point in declining them. Such a decline strikes me as Don Quixotic.

   Cf., advocates of more progressive income taxes are not expected to contribute the current difference to the state.

3. For not being a good speaker (e.g., unclear pronunciation, undisciplined, moody, PP-challenged, uses dense slides as an anchor).

3 Part 1 (non-technical)

Advice to *some* aspiring scientists and critique of *some* privileged scientists.

> Having turned sixty in February, I now belong to the elders. So I may be pardoned for promoting a good old vision, and arguing that it is still relevant, and maybe even more relevant than ever before.

4 Advice to SOME aspiring (i.e., non-tenured) scientists

I say *some*, because I don't really believe in generic advice. A good advice should be tailored to the personality of the advisee. So I guess the following advice is suitable for people who are somewhat similar to me, at least in some relevant aspects. In short, *you should take the following advice only if you feel that it empowers you; otherwise ignore it.*

One piece of advice: Don't be "humbled" by past research and researchers.

- On myself (Part 1): Proof of immodesty. (This is provided in order to establish credentials for the next item.)

  The proof: I think that I made important research contributions, of the calibre one can be proud of (and I'm indeed proud of them...).

- On myself (Part 2): I lack extraordinary skills of any type.

  In general, the gap between skills is grossly overstated. The difference between skills is quite minor if you take the correct perspective. Consider the skills that underlie:

  1. creating a scientific work (e.g., one presented in STOC'17),
  2. understanding such a work (when clearly presented),
  3. knowing a specific natural language (e.g., English),
  4. knowing no language.

  The gap between 1 and 2 pales when compared to the gap between 2 and 3, which pales when compared to the gap between 3 and 4.

  What may distinct me (and account for my contributions) is that I possess and am guided by good attitudes towards research.

- Good attitudes can be learn, adopted, and internalized.

- The main attitude: A commitment to Science.

  It is not a sacrifice (to some transcendental God) but a life choice.

  Submitting to it is being true to yourself (your choice). Betraying it for temporal benefits is pathetic.

Conclusion: Take yourself seriously. Adpot a serious attitude towards research.

5 Critique of SOME privileged (i.e., tenured) scientists

I say *some*, because only some deserve this critique.

- Failing to materialize the commitment to science. E.g., superficial reviews and evaluations.

  While the unprivileged may invoke the *defense of necessity* such a claim is to be rejected when made by privileged scientists (e.g., lack of an actual and specific existential threat).

- The duty of the privileged: Serve the scientific aspirations (esp., of the less privileged).

- Superficial reviews: Some underlying bad attitudes (where review is one service to the community at large):

  - Insisting on applications.

    Applications to practice: Hey, this is the theory of computation. We care about practice and we believe that our work may eventually affect computer practice, but our focus is on theory.

    Applications to theory: Work on interactive proof is required to have applications to PCP, work on PCP is required to have applications to inapproximability, etc. Sure, it is nice to have connections between different aspects, but a study of one natural aspect of the theory of computing is justified regardless of its applicability to another such aspect.
  - Referring to what is missing rather to what is present. A archetypical review comment is "It would have been nicer if the authors proved a stronger result."
  - Confusing what is obvious *a posteriori* (i.e., after understanding the new work) with what may be obvious *a priori*.
  - Complaining that "proofs are elementary, and have no new ideas." (Note the common relating of the two. How can a new work have no new ideas?)

  In general, not listening, rushing to grade.

- Superficial evaluations: Reduction to contents-oblivious measures.

6 Careerism vs the Vocation of Science (clarification)

It all boils down to the tension between careerism (i.e., a focus on personal benefits obtained via research) and the vocation of science, as articulated by Weber. Nice interpretations of these terms are implicit in MacIntyre (see "external goods"); the priority of the vocation follows from Kant's general imperative (see "means" vs "end").

N.B.: This priority does *not* call for totally ignoring career considerations, and avoiding external goods; it rather advocates not letting them dominate at all times. See the Kant's phrasing of "never *merely* as a means".

(Start with Weber, then MacIntyre, and end with Kant.)

- Kant (2nd formulation of the categorical imperative, 1785).

  Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end.

  (Readers typically omit the "own person" clause, but I want to stress it, because it is most relevant here, and because it is most profound in general.)

  (Of course, my point is that careerism and setting external goods as a goal implies (or is at least correlated) with using oneself as means.)

- Max Weber (*Science as a Vocation*, 1919).

  While focusing on one's career may generate some temporal benefits, it cannot generate true Science.

  Thus, in the long term, *careerism will even fail to serve the careerist.* Needless to say, it will always fail to give a feeling of responding to the vocation of Science, which means that *the careerist is doomed to a meaningless pursuit.* Such a person will be better off selecting less frustrating careers, since a "scientific career" is worthy its frustration only when the vocation of Science is present in it.

- MacIntyre (*After Virtue*, Chap. 14 ("the nature of the virtues"), 1981).

  External goods = *contingently attached* to the research activity by social circumstance; such goods include prestige, status and money. *Can be obtain in alternative ways.*

  Internal goods = can be *obtained only by performing research*, can only be *specified in terms of the research activity itself* and by means of examples from this activity, and can only be identified and recognized by the experience of participating in the activity in question.

  When external goods are achieved, they are always some *individual's property and possession.* Moreover, typically, the more someone has of them, the less there is for others. Hence, they are typically objects of competition in which there must be losers as well as winners.

  Internal goods are indeed the outcome of competition to excel, but it is characteristic of them that their achievement is a *good for the whole community* that participates in the activity.

N.B.: *The careerists treat themselves as means.*

Note: I quote MacIntyre, although his focus is on a moral theory, since he gives nice definitions of the external goods (which are the aim of the careerist) and the internal goods (which correspond to the vocation). (I have specialized his general treatment that refers to any social practice, which he defines in a way that clearly includes scientific research.)

7 Surviving bad times: The community and Society at large

I wish to stress that the issue reflected in "Careerism vs the Vocation of Science" is not specific to ToC. It is just one aspect of the Zeitgeist.

- Vulgar individualism (vs human as a social creature).

  By vulgar individualism I mean the fantasy of *fully autonomous* humans (cf., "There is no society, only individuals"), which leads to the fantasy of the Ubermensch.

- Elimination of meaning (vs the need for it [V. Frankl]).[1]

There is a connection: Meaning is based on interaction, on a social context (e.g., a research community).

I propose that a small community (like the ToC community) can resist the fate of society at large and create an island of meaning and social solidarity. This is possible in a scientific discipline because its core refers to meaning and to communication.

(this was the vision...)

[1] See also *Escape from Freedom* [Erich Fromm].

8 Part 2 (technical)

On Doubly-Efficient Interactive Proof Systems.

9 Interactive Proof Systems [Goldwasser-Micali-Rackoff]

The claim is "I love you (and you love me)" Petra and Camille (not Venus) interact throughout the movie in order to prove it. A short extract of their interaction reads:

Petra: What would we do, as friends?

Camille: Have fun.

Petra: Fun – sounds like a buddy movie.

Camille: Yes, exactly. Like Thelma and Louise. But... without the guns.

Petra: Oh, well, no guns, I don't know...

When Night is Falling (Canada, 1995).

10 Interactive Proof Systems [GMR]

In the actual definition, the verifier is PPT, whereas the prover is computationally unbounded. They interact with the intension of convincing the verifier that the common input x is in a predetermined set.

Completeness: If x is in the set, then the verifier always accepts (under a suitable prover strategy)

Soundness: If x is not in the set, then the verifier reject w.p. at least 1/2, no matter what strategy the (cheating) prover employs.

The celebrated theorem of [Lund-Fortnow-Karloff-Nisan and Shamir], asserting that every set in $\mathcal{PSPACE}$ has an interactive proof system, has overshadowed the notion of interactive proof systems and the existence of natural questions to be studied wrt it.

In particular, Petra is not computationally unbounded; so how can interactive proof systems help her?

11 Doubly Efficient Interactive Proof Systems [Goldwasser-Kalai-Rothblum]

Here the *prescribed prover* is PPT. Hence, the proof system can only be used to prove membership in sets that are in $\mathcal{BPP}$, so *what's the point*? The point is that the verifier may be more restricted; in particular, the complexity of verifying may be lower than the complexity of deciding. For concreteness and simplicity, we consider verifiers that run in almost linear time. The resulting proof systems are called doubly efficient.

We stress that soundness should still hold with respect to computationally unbounded cheating provers. Only the good/honest/prescribed prover is in PPT, whereas "security" (i.e., soundness) is information theoretic (i.e. statistical). In particular, no computational assumptions are used (i.e., the results are "unconditional").

> You will probably hear about computationally-sound systems at the Friday workshop on IP/PCP.

Historical note: Such interactive proof systems were called "interactive proofs for muggles" [GKR] and "delegation schemes" [RRR]. The latter term confused the archetypical application (verifying the correctness of a claimed computational outcome) with the notion.

12 Simple Doubly Efficient Interactive Proof Systems

Ex 1: In some cases, *almost-linear-time verifiable NP-witnesses can be found in polynomial-time* (e.g., perfect matching, primality, t-Clique for constant $t \geq 3$).

Ex 2 [GR17]: $\texttt{no-}t\texttt{-Clique}$, for constant $t \geq 3$. Representing the input graph by its adjacency matrix $A = (a_{i,j})_{i,j \in [n]}$, with $a_{i,i} = 0$, we consider

$$\sum_{\{i_1,\ldots,i_t\} \in \binom{[n]}{t}} \prod_{j,k \in [t]} a_{i_j,i_k}.$$

We want to prove that this sum equals zero.

For $\ell = \log_2 n$, let $[n] \equiv \{0,1\}^\ell \subset F^\ell$, where F is a sufficiently large finite field. Then, consider

$$\sum_{\overline{w}_1,\ldots,\overline{w}_t \in \{0,1\}^\ell} \prod_{j,k \in [t]} \sum_{(u,v) \in E} \texttt{EQ}(\overline{w}_j \overline{w}_k, uv)$$

where $\texttt{EQ}(y,z) = \prod_i (y_i z_i + (1-y_i)(1-z_i))$. Points to stress include that the sum over F is over the integers (since $|F| > 2^\ell$ and each term in in $\{0,1\}$), and that $\texttt{EQ}$ is a low-degree polynomial.

Now, apply the *sum-check protocol* (to the outer sum). After $t\ell$ rounds, the verifier is left with evaluating an arithmetic expression with $t^2 \cdot |E|$ terms. Recall: The soundness error is $t\ell \cdot t^2/|F|$.

> **After thought**: Boaz Barak has suggested to start by saying that when presenting interactive proof systems for co$\mathcal{NP}$ we start with an arithmetization that has an exponential number of terms. This implies that the interactive proof has a linear number of rounds and that implementing the prover strategy requires exponential time (in order to evaluate all partial sums of the Arithmetic formula). In conrast, here we start with an arithmetization that has a polynomial number of terms; hence, the interactive proof has a logarithmic number of rounds and the prover strategy can be implemented in polynomial-time.

13 The Sum-Check Protocol [LFKN]

For $f : H^m \to \{0,1\}$ (e.g., $H = \{0,1\}$), let $\widehat{f} : F^m \to F$ be a polynomial (over the finite field $F \supset H$) of individual degree d that agrees with f on H^m.

Proving that $\sum_{\overline{\alpha} \in H^m} f(\overline{\alpha})$ equals v (where the arithmetics is of F) is equivalent to proving that

$$\sum_{\overline{\alpha} \in H^m} \widehat{f}(\overline{\alpha})$$

equals $v_0 \overset{\text{def}}{=} v$. The proof proceeds in iterations such that in the i^{th} iteration the prover sends $p_i(z) \overset{\text{def}}{=} p_{r_1,...,r_{i-1}}(z)$ such that

$$p_{r_1,...,r_{i-1}}(z) \overset{\text{def}}{=} \sum_{\alpha_{i+1},...,\alpha_m \in H} \widehat{f}(r_1, ..., r_{i-1}, z, \alpha_{i+1}, ..., \alpha_m)$$

and the verifier checks whether $\sum_{\alpha \in H} p_i(\alpha) = v_{i-1}$ and rejects if inequality holds. Otherwise, the verifier replies with a random $r_i \in F$, and both parties set $v_i = p_i(r_i)$. After all iterations are completed, the verifier checks whether $\widehat{f}(r_1, ..., r_m) = v_m$. Hence, the verifier needs only evaluate $\widehat{f}$ at a single point.

Analysis: The foregoing protocol has soundness error $m \cdot d/|F|$, since each p_i has degree d.

14 Simple Doubly Efficient Interactive Proof Systems for Locally Char. Sets [GR17]

Def: S is locally characterizable if for a logarithmic function ℓ (i.e., $\ell(n) = O(\log n)$) and uniform Boolean formulas $\phi_n : \{0,1\}^{\ell(n)+m(n)} \to \{0,1\}$ and $\pi_n : \{0,1\}^{\ell(n)} \to [n]^{m(n)}$ of poly($\ell(n)$)-size it holds that $x \in S$ if and only if for every $w \in \{0,1\}^{\ell(n)}$

$$\Phi_x(w) \overset{\text{def}}{=} \phi_n(w, x_{\pi_n(w)})$$

equals 0, where $n = |x|$ and $x_{\{i_1,...,i_m\}} = (x_{i_1}, ..., x_{i_m})$.

For example, no-t-clique and no-t-sum.

Thm: Every locally characterizable set has a doubly efficient interactive proof system (with prover's time $2^{\ell(n)} \cdot \widetilde{O}(n)$, verification time $\widetilde{O}(n)$, and $\ell(n)$ rounds).[2]

Proof Idea: For a finite field F greater than $2^{\ell(n)}$, let $\widehat{\phi}_n$ and $\widehat{\pi}_n$ be low degree "extensions" of ϕ_n and π_n, resp. Then, we wish to prove that

$$\sum_{w \in \{0,1\}^{\ell(n)}} \widehat{\phi}_n \left(w, \left(\sum_{\alpha \in \{0,1\}^{\log n}} \texttt{EQ}(\widehat{\pi}_n(w)_i, \alpha) \cdot x_\alpha \right)_{i \in [m(n)]} \right)$$

equals zero.

[2] This is one instantiation of a more general result. Furthermore, the system is of the public coin type.

$\widehat{\phi}_n(\cdot, \mathtt{EQ}(\cdot,\cdot))$ is a low degree polynomial.

After applying the Sum-Check Protocol, the verifier needs to evaluate $\widehat{\phi}_n$ at one point, and the dominant computation is of the internal sum (over n terms).

Note: For every $w \in \{0,1\}^{\ell(n)}$, it holds that $\sum_{\alpha\in\{0,1\}^{\log n}} \mathtt{EQ}(\widehat{\pi}_n(w)_i, \alpha) \cdot x_\alpha$ equals $x_{\pi_n(w)_i}$.

After thought: In the STOC presentation, I rushed through the following three slides, presenting and discussing only the statements of the results.

15 Doubly Efficient Interactive Proof Systems for $\mathcal{NC}$ [GKR]

Thm: Every set in log-space uniform $\mathcal{NC}$ has a doubly efficient interactive proof system. For depth $d(n)$, we use $d(n) \cdot \text{poly}(\log n)$ rounds.[3]

Proof Idea: For $d = d(n)$ and $k = k(n) = \text{poly}(n)$, consider the intermediate values of the computation of $C_n(x)$. Letting $\alpha_i \in \{0,1\}^k$ describe the values assigned to layer i, we wish to verify that $\alpha_d = x0^{k-n}$, that the α_i's are consistent with the gates of C_n, and that $\alpha_0 = C_n(x)0^{k-1}$. Considering low-degree extensions of the α_i's, denoted $\widehat{\alpha}_i : F^m \to F$, and letting $\psi_i : H^{3m} \to \{0,1\}$ describe the circuit, we wish to verify for every $i \in [d]$,

$$\widehat{\alpha}_{i-1}(\overline{z}) = \sum_{\overline{u}\in H^m} \mathtt{EQ}(\overline{z},\overline{u}) \cdot \sum_{\overline{v},\overline{w}\in H^m} \widehat{\psi}_i(\overline{u},\overline{v},\overline{w}) \cdot (1 - \widehat{\alpha}_i(\overline{v}) \cdot \widehat{\alpha}_i(\overline{w}))$$

(for an $\mathtt{NAND}$-gate).

Note that we can easily verify that $\widehat{\alpha}_d$ and $\widehat{\alpha}_0$ are low degree extensions of $\alpha_d = x0^{k-n}$ and $\alpha_0 = C_n(x)0^{k-1}$, resp. We reduce the verification of the value of $\widehat{\alpha}_{i-1}$ at a (random) point to the verification of the value of the value of $\widehat{\alpha}_i$ at a new random point, by using the Sum-Check Protocol. We, however, face two problems:

1. Within the Sum-Check Protocol, the verifier needs to evaluate $\widehat{\psi}_i$ in almost linear-time, whereas the straightforward computation is almost linear in $|H|^{3m} = k^3 = \text{poly}(n)$.

   Can be solved in case C_n is "highly uniform" (i.e., the Boolean function ψ_i has a small uniform formula, where small means $\text{poly}(m \log |H|) = \text{poly}(\log n)$). (This condition may not hold for C_n itself, but it holds for a circuit that first constructs C_n and then uses a universal (evaluation) circuit.)

2. After the Sum-Check Protocol, the verifier needs to verify the value of $\widehat{\alpha}_i$ at two points (rather than at one).

   Solved by checking the value at a random point on a line that connects these two points.

[3] This is one instantiation of a more general result. Furthermore, the system is of the public coin type. Actually, the round complexity is $O(\log n)^2 \cdot d(n)$.

16 Doubly Efficient Interactive Proof Systems for $\mathcal{SC}$ [Reingold-Rothblum-Rothblum]

Every set in $\mathcal{SC}$ has a doubly efficient interactive proof system of constant number of rounds. Extends to $\mathrm{TiSp}(\mathrm{poly}, n^{0.5})$.

Although Allan Borodin has taught us four decades ago that circuit depth is related to TM space, this relation is not tight enough for having one of these results (i.e., GKR and RRR) supersede the other. Also note that the result of RRR (which extends to randomized $\mathrm{TiSp}(\mathrm{poly}, n^{0.5})$) is almost optimal, since doubly-efficient interactive proof systems exist only for sets in the intersection of $\mathcal{BPP}$ and almost linear space. This does leave a gap between $\sqrt{n}$ and n, and more importantly between simultaneous time-space and the intersection of time and space.

After thought: Avi and Boaz asked if I have a conjecture as to where the truth lies (w.r.t the latter gap) and whether the class of sets having doubly-efficient interactive proof systems is closed under complementation. I was too absentminded to answer, but I care much more about the first question. In a later private discussion with Ron Rothblum, he noticed that *if the said class equals one of the two extremes* (i.e., randomized $\mathrm{TiSp}(\mathrm{poly}, n^{0.5})$ or $\mathcal{BPP} \cap \mathrm{Dspace}(\widetilde{O}(n))$), *then it is closed under complementation* (since each of the extremes is closed in that sense).

One of the proof ideas: *Batch verification*; that is, verifying t claims at a cost lower than running the original verification procedure for t times.

- A sanity check: Possible (via IP=PSACE) if one does not care of the prover's time.
- A toy example: Batch amplification for UP (i.e., NP with unique witnesses). The verification time and communication for $t(n)$ instances is $t(n)^{1/O(1)} \cdot c(n) + t(n)$, where $c(n)$ denotes the complexity for a single instance.

  This is a toy only when compared to what is actually needed, which is batch verification for interactive proof systems. We need an adequate notion of uniqueness (called *unambiguous*) as well as a PCP-version of IP (which supports querying bits in the communicated messages rather than reading the entire messages).

17 Batch verification for UP [RRR]

Standard PCP systems use "input-oblivious queries" (i.e., the query distribution is independ of the specific input, although the final decision depends on it). Also, the NP-witness is easy to extract from a valid PCP-oracle (e.g., it can be a prefix of it), and the oracle length $c(n)$ may be almost linear in the witness length. For any d (e.g., $d = \sqrt{t(n)}$)), we use a parity-check function $F : \{0,1\}^{t(n)} \to \{0,1\}^{O(d \log t(n))}$ such that for every preimage y and image v, the Hamming ball of radius d centered at y contains at most one preimage of v.

1. Let π_i be the PCP-oracle derived from the unique NP-witness of the i^{th} input. The prover constructs a matrix with rows π_i, and computes the party-check of each column. It sends these $c(n)$ parity-check values to the verifier.

   (Hence, communication is approximately $d \cdot c(n)$.)

2. The verifier selects queries $j_1, ..., j_q \in [c(n)]$, and sends them to the prover, who responds with the corresponding columns.

   (Hence, communication is $q \cdot t(n)$.)

   The verifier checks that the PCP-verifier accepts in all rows, and that the the revealed columns satisfy the corresponding parity-checks.

3. The verifier selects $O(t(n)/d)$ random rows, sends their indices to the prover, who responds with the corresponding rows.

   (Hence, communication is $O(t(n)/d) \cdot c(n)$.)

   The verifier checks that the revealed rows are legal (actually the unique) PCP-oracles for the corresponding inputs, and that they match the revealed columns.

The basic intuition is that Step 1 leaves the prover with the option of either cheating in Step 2 on more than d entries of some column or providing the "correct" values of these columns (where a correct value for a wrong input is defined as 0). In the first case, it will be detected (whp) by Step 3, and otherwise an input not in the set will be rejected (whp) by Step 2.

18 Wider perspective: the source of proofs

Proofs do not fall out of thin air.

Setting 1: The prover's on-line work (i.e., doubly efficient interactive proof systems, reviewed here).

Setting 2: Provided by the (higher level) application/user + PPT.

E.g., an NP-proof provided as auxiliary-input to a zero-knowledge interactive proof system.

Setting 3: Constructed in PPT with oracle access to deciding.

Three *different notions of relatively efficient provers*, which deserve further study. We focus on statistical soundness, but computational-soundness is also interesting (cf., [Kilian92, Micali94] and SZK arguments).

Historical note: See Sec. 2.1.3 ("How efficient should the prover be") of *A Taxonomy of Proof Systems*, 1995.

19 END

Slides available at `http://www.wisdom.weizmann.ac.il/~oded/T/k-lecture.pptx`

Related papers and expositions available at `http://www.wisdom.weizmann.ac.il/~oded/de-ip.html`