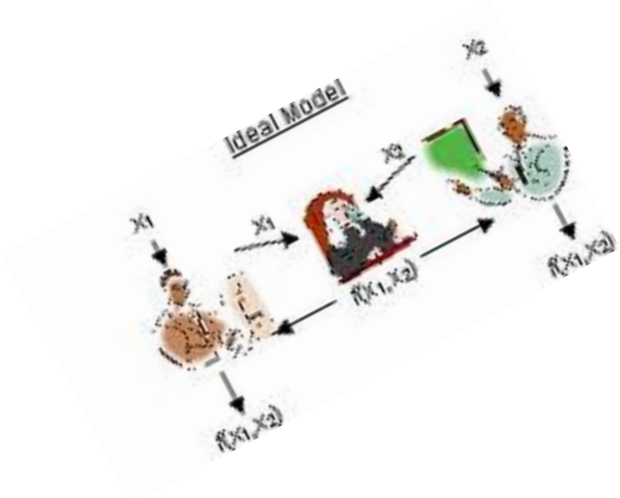
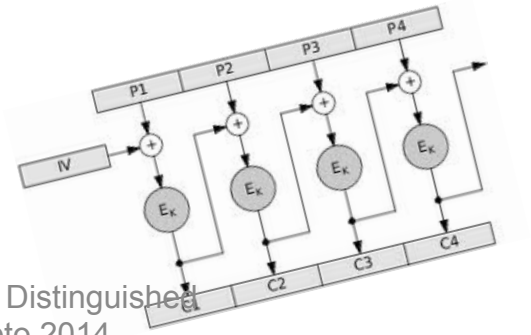
Caught Between Theory, Practice and Peer Review

Mihir Bellare

UCSD



Regrettably, your submission was not accepted to … We received many outstanding submissions, and the selection of which ones to include in the program was not an easy task …



values, tastes, judgments, …

Disciplinary culture

Papers

Theory versus practice

Peer review

Affect our
success on job market, promotions, motivations, choice of problems, expository style, self-image, opinions of others, community impact, …

change

understanding

Sociology, psychology and guesswork

Kahneman & Tversky
Biases and their role in decision making

Kuhn
The nature of normal science

Today

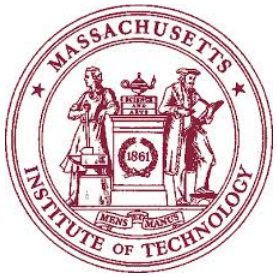
Theory versus
practice

Peer review

Anecdote, discussion, cultural phenomena, possible explanations

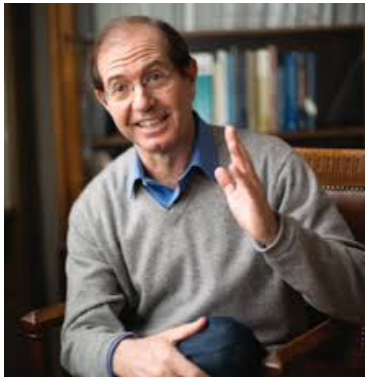
Part I: Theory vs. Practice

A Tale of Two Cultures



MIT, 1987

6.875 Cryptography and Cryptanalysis



Pseudorandom bit generators [BM,Y]
Pseudorandom functions [GGM]
Probabilistic encryption, semantic security [GM]
Digital signatures unforgeable under adaptive chosen-message attack [GMRi]
Zero-knowledge interactive proofs [GMRa]

Foundations

that are important to good **practice**

What attracted me:

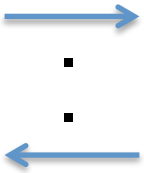
Cryptography = Philosophy made precise

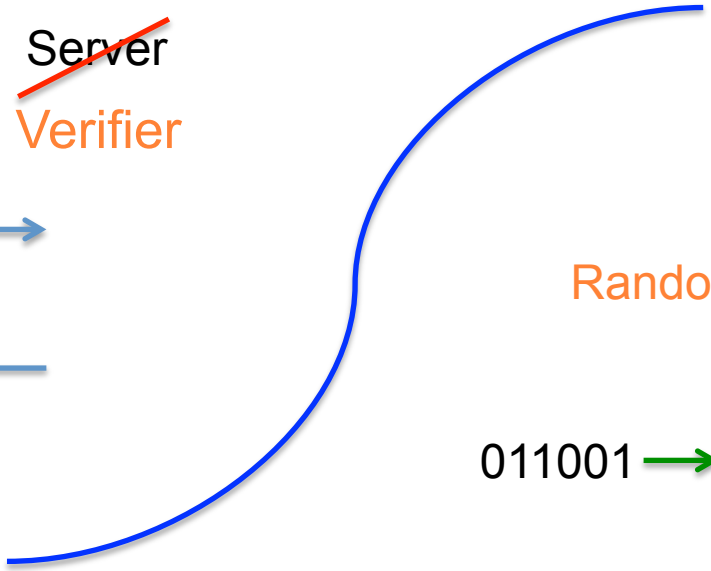
Security in an imaginative context

Humanist perspective

Client Server

Prover Verifier





Encryption

Randomness vs computation

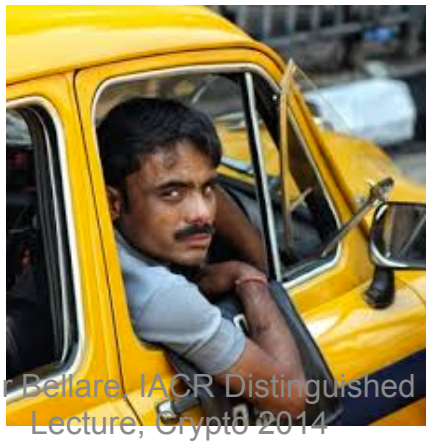
011001 → PRG → 10011000101101

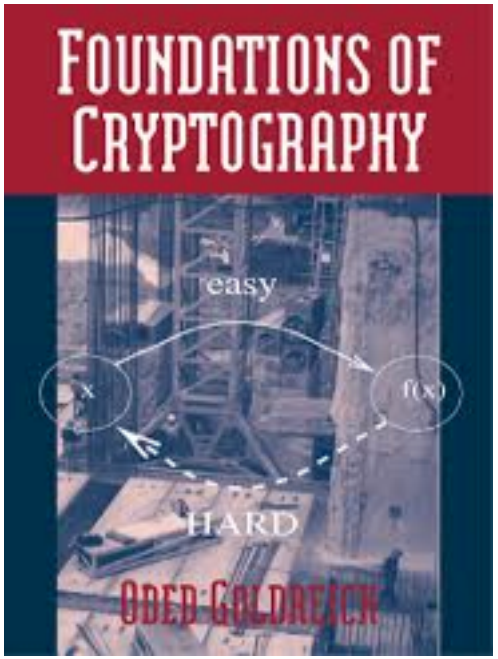
A Way of Life


Amarone
zero-knowledge
Verdi
Caruso
Benigni
Fellini
sushi
Masada
semantic-security
goat
espresso
Madonna

A Way of Life

Spoke particularly to me, who had come to science lately, my first interests being literature and history



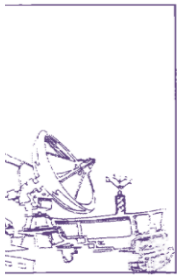


The Philosophic Culture of Cryptography

Humanist motivations
Strong definitions of security
Proofs by reduction
Asymptotic analysis
Assumption minimization
Algebraic starting points
In-principle achievability

Typical Theorem: If one-way functions exist, then there exists a S-secure scheme for goal G.

The Philosophic Culture: The adversary's perspective



NATIONAL SECURITY AGENCY

CRYPTOLOG

Humanist motivations
Strong definitions of security
Proofs by reduction
Asymptotic analsis
Assumption minimization
Algebraic starting points
In-principle achievability s

"Those of you who know my prejudice against the ``zero-knowledge" wing of the philosophical camp will ..."

"Don Beaver ... a spell-binding, charismatic preacher ... has captured from Silvio Micali the leadership of the philosophic wing of the US East Coast"

"Even if his results are correct ... it may be good statistics (or mathematics, or computer science or philosophy) but it is not good cryptanalysis ..."

Whenever I suggest to do something practical, one of you jumps out the window and the other out the door!

Whenever I suggest to do something practical, one of you jumps out the window and the other out the door!

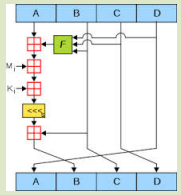
+ IBM =

Interview!

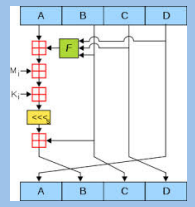
IBM Research

Don Coppersmith

So what's new with hash functions?



MD4, MD5



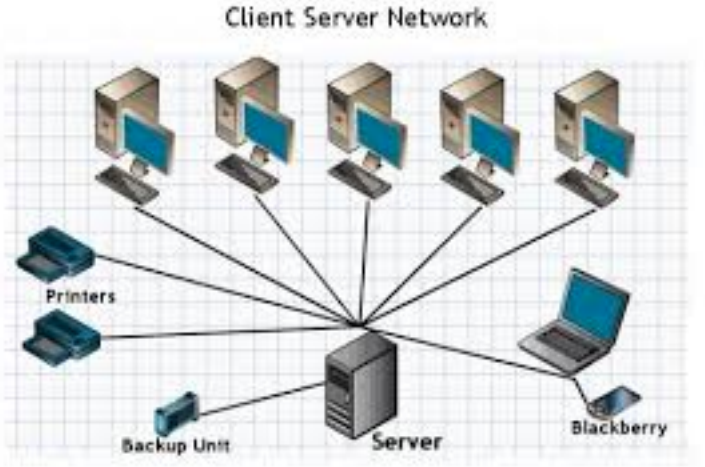
MD5

MD4/MD5 were amongst the most influential pieces of practical cryptography of their decade. Ubiquitously used, 720 places in Microsoft Windows alone.

Verdict

IBM Research

plaNET


Client Server Network
Printers
Backup Unit
Server
Blackberry


ACCEPTED

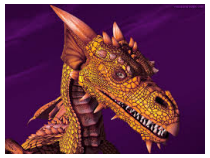
Phil Rogaway

DES MD5

Kerberos CBC

MAC PKCS#1

SHA1



Phil Rogaway

DES MD5

Kerberos CBC

MAC PKCS#1

SHA1

Theory

- Definitions of security ✔
- Confidence via proof ✔
- Algebraic starting points
- Asymptotic security
- Public-key cryptography ✔
- MPC, ZK, OT, …

Practice

- Informal security requirements
- Confidence via cryptanalysis
- Confusion-diffusion starting points ✔
- Concrete security ✔
- Symmetric cryptography ✔
- Session-key distribution, … ✔

Practice-oriented provable security

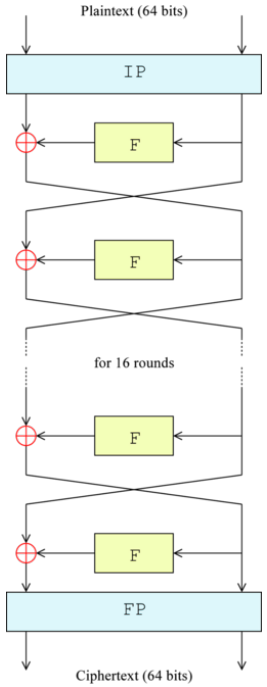
``An apparently arbitrary element, compounded of personal and historical accident, is always a formative ingredient of the beliefs espoused by a given scientific community at a given time.'' **Kuhn**, *Structure of Scientific Revolutions*.

DES: What I had heard at MIT …

"Some sort of engineering-based one-way function … "

Not science

Not even right

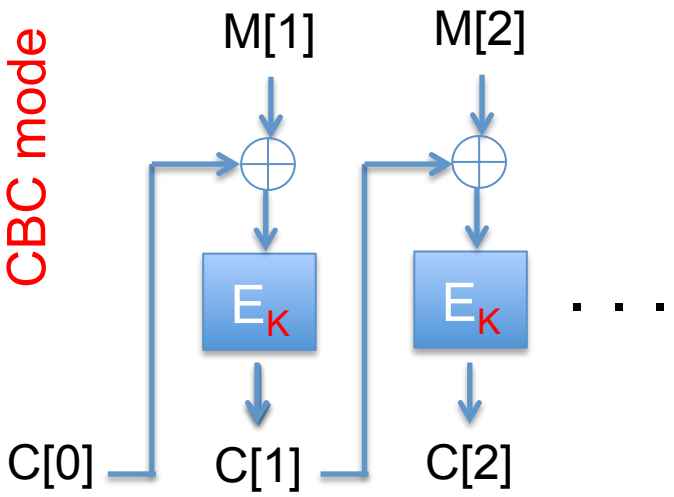






PRFs

PRPs



We modeled blockciphers as "finite" PRPs / PRFs

Confusion-Diffusion constructs become base primitives whose assumed security can be used to validate higher-level constructs.

Thm: [BDJR98]

Let $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a blockcipher.
Let SE be the CBC symmetric encryption scheme based on E.
Suppose messages are m blocks long.
Let A be a time t ind-cpa adversary against SE.
Then we can construct a time t prp-adversary B against E such that

$$\mathrm{Adv}^{\text{ind-cpa}}_{SE}(A) \le 2\,\mathrm{Adv}^{\text{prp}}_{E}(B) + \frac{2q^2m^2}{2^n}$$

Gave birth to provably-secure symmetric cryptography:

- Proofs of existing modes
- New modes
- New goals: authenticated encryption, format-preserving encryption, …

Advantage functions.

Thm: [BDJR98]

Let $E : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a blockcipher.
Let SE be the CBC symmetric encryption scheme based on E.
Suppose messages are m blocks long.
Let A be a time t ind-cpa adversary against SE.
Then we can construct a time t prp-adversary B against E such that

$$\mathrm{Adv}^{\text{ind-cpa}}_{SE}(A) \le 2\,\mathrm{Adv}^{\text{prp}}_{E}(B) + \frac{2q^2m^2}{2^n}$$

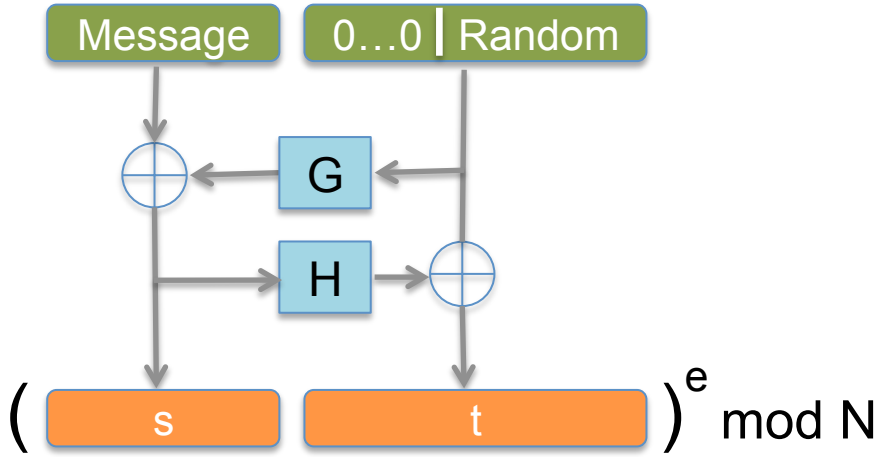
Confusion-diffusion constructs have strengths beyond those captured by existing formal definitions

Random-oracle model [BR93a]

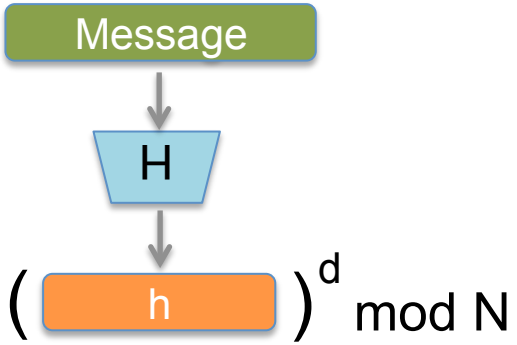
Scheme algorithms and adversary have oracle access to

H(x)
If T[x] is undefined then pick T[x] at random
Return T[x]

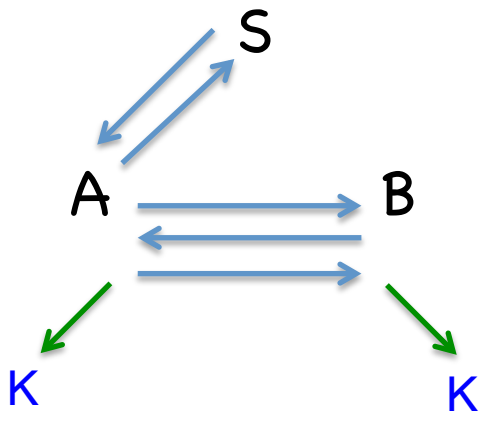
OAEP public-key encryption scheme [BR94]



Full-domain-hash (FDH) signatures and PSS [BR93a,BR96]

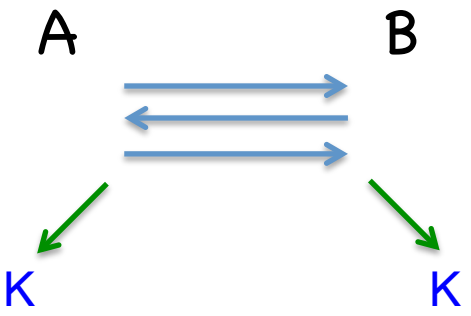


Session-key distribution [BR93b,BR95,BPR00]



Three-party setting
Kerberos, …

Two-party setting
IKE, …



Session key K must be authentic, private and fresh.

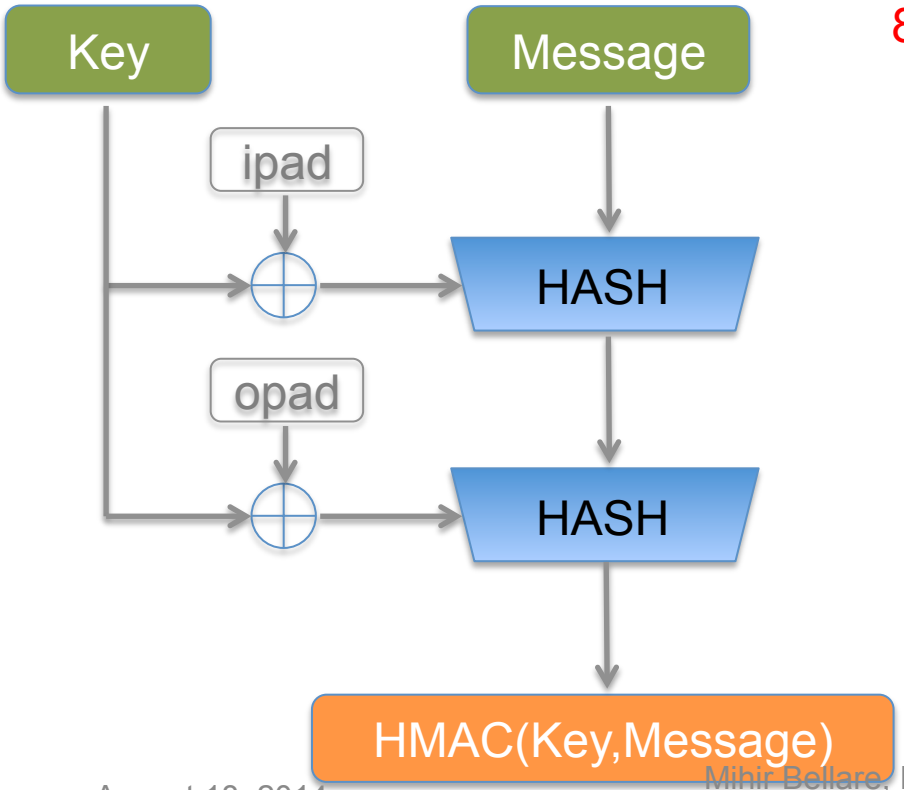
Harder than it looks …

We gave definitions and proven-secure protocols

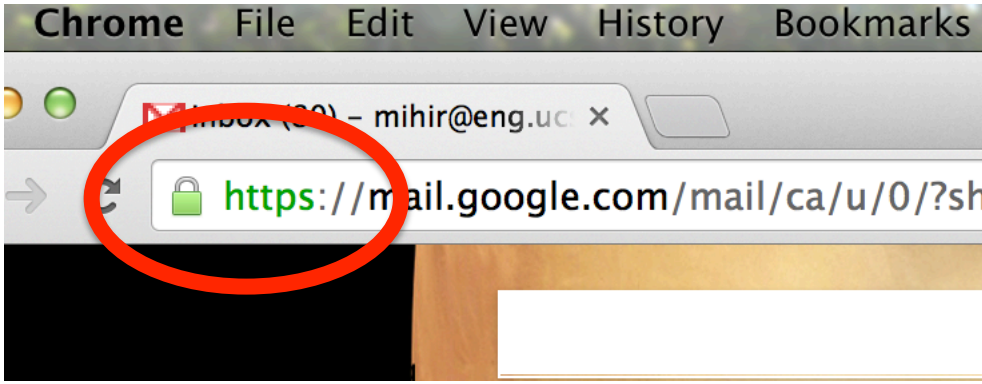
How can we authenticate messages with hash functions (like MD5) rather than with blockciphers (like DES)?

Ran Canetti Hugo Krawczyk

HMAC [BCK96]



HMAC is in TLS, IPSEC, SSH, IEEE 802.11e, …



Impact

- Over 40 standards based on this line of work
- Changed perception of theory

HMAC [BCK96] – RFC 2104, ANSI X9.71, NIST FIPS 198, IEEE 802.11
OAEP [BR94] – RSA PKCS#1 v2.1, ANSI X9.44, CRYPTREC, ISO/IEC 18033-2, RFC 3447, RFC 3560
PSS [BR96] – RSA PKCS#1 v2.1, ANSI X9.31, CRYPTREC, IEEE P1363a, ISO/IEC 9796-2, NESSIE, RFC 3447
OCB [RBBK01] – RFC 7253, ISO/IEC 19772
FFX [BRS10] – NIST-800 38G
DHIES [ABR01] – ANSI X9.63, IEEE P1363a, ISO/IEC 18033-2, SEC
EAX [BPW04] – ANSI C12.22, ISO/IEC 19772
…

Nowadays standards bodies expect proofs for higher-level constructs.

Practical crypto ≠ Real-world security

Doesn't address:

- Implementation error
- Side-channel attacks
- Insider attacks
- PRISM, XKEYSCORE, BULLRUN, MUSCULAR, LUSTRE, …
- …

``Encryption works. Properly implemented strong cryptosystems are one of the few things that you can rely on.'' **Edward Snowden**.

Retrospective: Utility of theory

The most useful thing theory has to offer practice is DEFINITIONS.

NOT efficiency improvements to theoretical schemes.

Retrospective: The philosophic culture

Narrow

despite apparent breadth

Hamlet: There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.

Confusion-diffusion primitives
Practical motivations
Formal methods
…

The hardest task for the MIT graduate is to unlearn …

In the company of theoreticians

I feel like a practioner

In the company of practioners

I feel like a theoretician

It is not just me …

Our research community is caught between theory and practice

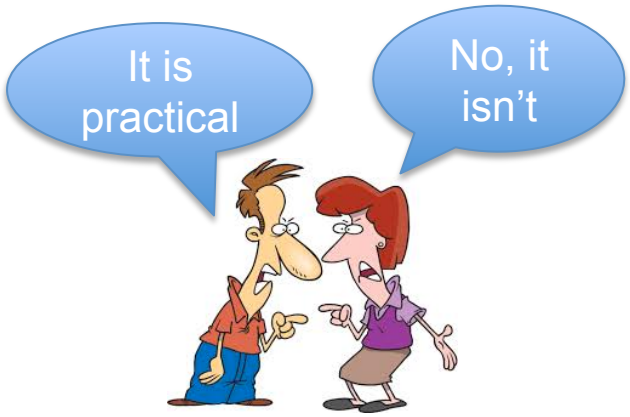
THEORY

PRACTICE

Symptoms of being caught-in-between

exaggeration
sensitivity
contention

politics



Most theory papers claim practical applications or motivations

But practioners say almost none of these papers actually delivers anything of practical utility

A lot of work is about efficiency improvement

But for primitives that are utility-free

Meanwhile many real practical problems are not even being addressed.

When different people say ``practical'' they mean different things

Needed:

Definitions

Foundations

Defining ``Practical”

The best notion

UTILITY: X is USEFUL

Lots of people use it and want it.
It has a market. It has social value.
It solves a problem people actually want solved.
It makes us more secure in real life.

MONEY: people PAY MONEY for X

A for-profit entity buys it.
Individuals pay for it. We have a customer.

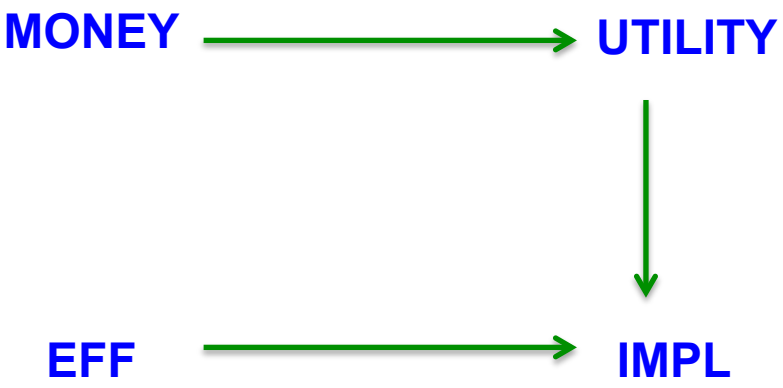
IMPL: we IMPLEMENTED X

I wrote, or got someone to write, code for it.

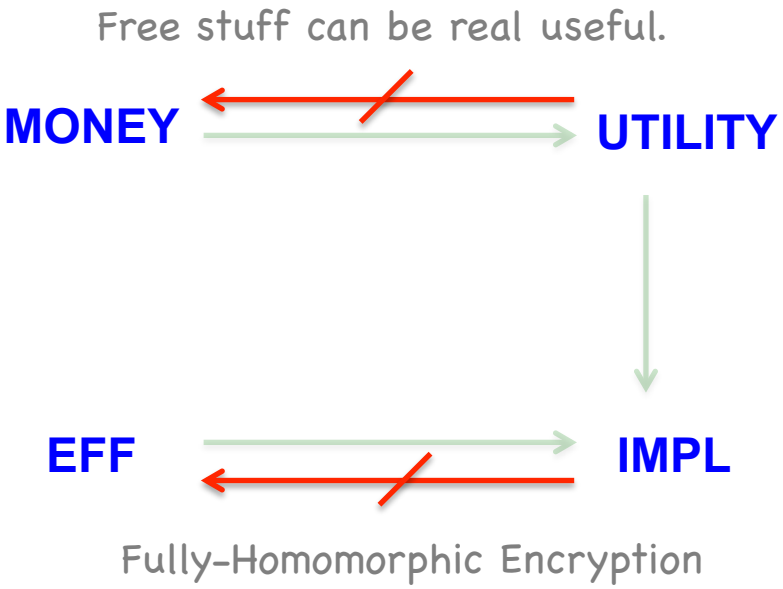
EFF: X is EFFICIENT

Less than 100 group operations? No NIZKs?
Cycles per byte?

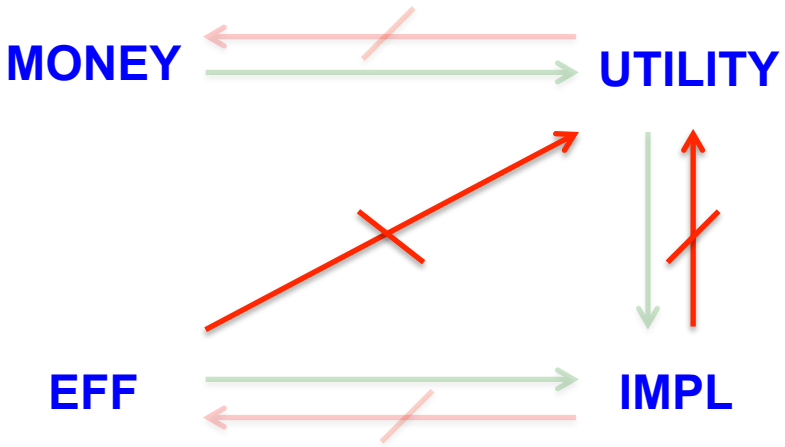
Relations between notions of practicality



Relations between notions of practicality



Relations between notions of practicality



Almost everything is a separating example.

You can make your primitive as fast as Usain Bolt, but it doesn't help if nobody wants it.

Towards achieving utility

Bottom-up works better than Top-down

Solution = ?

Real Problem

Favorite primitive

Application = ?

Founding Cryptography on Oblivious Transfer

Joe Kilian, MIT

STOC 88

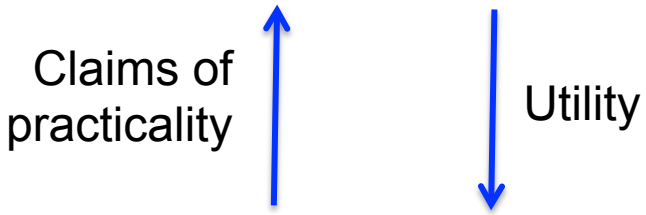
Introduction

Cryptographers seldom sleep well at night [M] … A poly-time algorithm for factoring would certainly prove more crushing than any paltry fluctuation of the Dow Jones …

References

[M] Micali, Silvio, Personal Communication.

Number of occurrences of word “practical” in [BlMi84,GM84,GGM86,GoMiRa89] : **0**

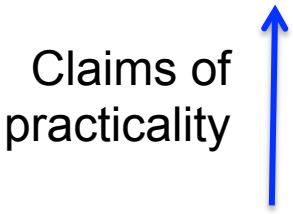


Why?

True (internal) motivation ≠ Stated (external) motivation

Technical challenge
Philosophic interest

Practicality



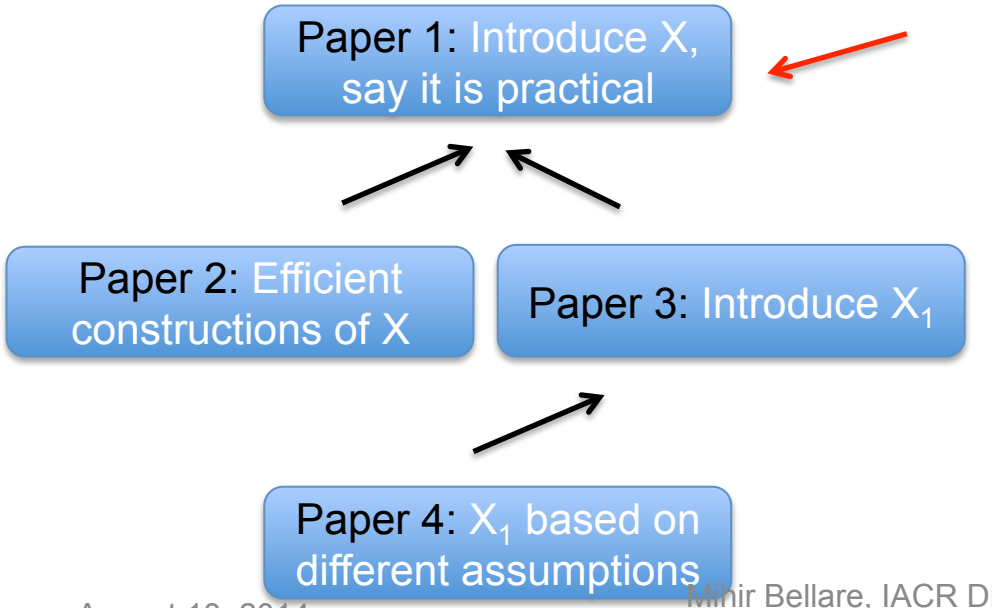
Why?

Pressure: get papers accepted, get grants funded, get jobs?

Claims of practicality

Why?

A genuine belief in practicality fostered by delegated motivation



But X never was of genuine practical utility

Body of work whose practicality is justified by citation to Paper 1

Such bodies of work can be large and long-lived

Claims of practicality

Why?

A genuine belief in practicality

fostered by delegated motivation

and by peer review

My paper was rejected because reviewer said it had no practical applications. It's those !*& practioners!

IT - Important Theoretician

Not true!
Ironically, it was a theoretician.

Part II: Peer Review

TCC

CRYPTO

EUROCRYPT

ASIACRYPT

PKC

. . .

Program Committee (PC) decisions → This is our life …

- Personally
- As a community

Affect our choice of problems, expository style, field trajectory, confidence, impact.

How well does the process work?

Not very well …

``We portray peer review to the public as a quasi-sacred process that helps to make science our most objective truth teller. But we know that the system of peer review is biased, unjust, unaccountable, incomplete, …, often insulting, usually arrogant, occasionally foolish, and frequently wrong.''

Richard Horton, Editor, The Lancet, 2000.

``… peer review makes the ability to publish susceptible to control by elites and to personal jealousy … If you do not belong to this tight fraternity it becomes extremely difficult to gain a hearing for your work …''

Robert Higgs, Nature Magazine, 2007.

``… reviewers tend to be especially critical of c t their own views and lenient towards those that match t ze with the established experts' are more likely to see pri

paraphrasing Thomas Kuhn

Reviews may be biased, unjust, insulting, arrogant, foolish, wrong. Reviewers can be elitist, critical of conclusions that contradict their own, unaccountable and irresponsible.

Kübler-Ross model

Denial
Anger
Bargaining
Depression
Acceptance

NO!

How dare you suggest this!

How we feel about PC decisions, reviews and the process

Reviews may be biased, unjust, insulting, arrogant, foolish, wrong. Reviewers can be elitist, critical of conclusions that contradict their own, unaccountable and irresponsible.

Almost all authors complain.

PC members complain routinely.









We don't complain enough

Complaints are private.

They should be public.

Apathy has set in.

The reviews I got are wrong and biased. Follow-on work to mine by friends of PC members got accepted.

But …

That's how the system is, has been and always will be. There is nowhere to appeal or complain. Nothing to do but have a drink and forget.

Roots

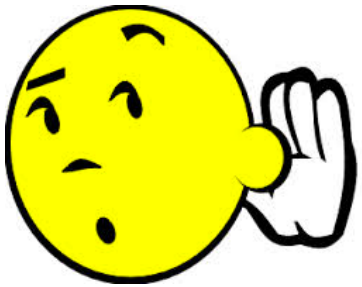
PC = The Adversary

My paper provides …

Cool! Is it published?

Rejected from Crypto …

Pr[accept] = 25%



Today: An attempt to understand peer review

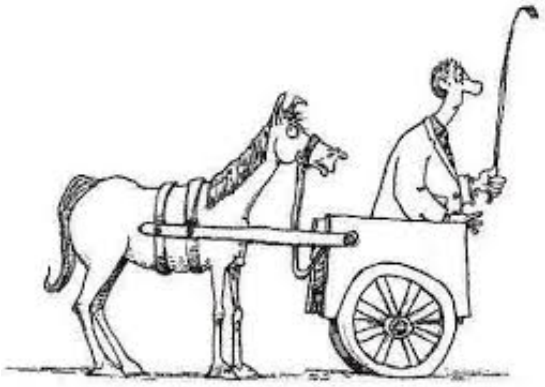
- Critique – Issues and phenomena
- Explain – Via sociology, psychology and guesswork
- Model – Peer review as a judicial system

A few clarifications

Other reviewing and publication systems being proposed in our community are subject to the same critiques since they continue to be based on peer review.

I am not exempt from any of my critiques.

NO, I don't have a solution. We benefit from understanding the problem first.



Denial
Anger
Bargaining
Depression

Obstacles to acceptance of the problem :

May think this happens.

Would like to think this happens

Reviews may be biased, unjust, insulting, arrogant, foolish, wrong. Reviewers can be elitist, critical of conclusions that contradict their own, unaccountable and irresponsible.

We, as reviewers and PC members, are unbiased, just, fair, accountable, responsible, polite, humble, wise and correct. There are no elites. We welcome views critical of our own.

Denial
Anger
Bargaining
Depression

Obstacles to acceptance of the problem :

Bad things happen only when bad people are involved.

But I, my friends, and most of us, are fundamentally good people.

Bad things happen only when bad people are involved.

This viewpoint is in contradiction with

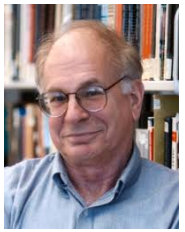
- Accepted understanding in modern psychology and sociology
- The history of our species

Powerful influence on decision theory and decision making in many domains

Judgment under Uncertainity: Heuristics and Biases. Science, 1974. 30,100 citations.

Kahneman Tversky

Nobel Prize in Economics, 2002



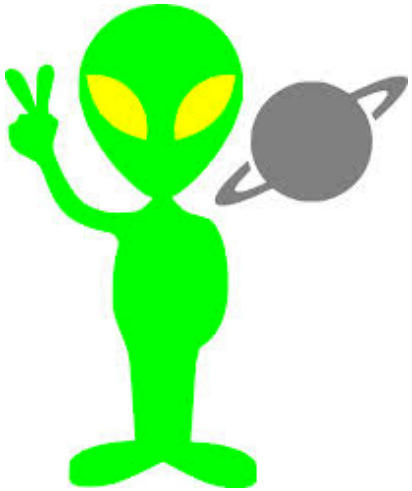
Biases are ubiquitous, well-studied and documented.

Anchoring – Value scale influenced by one distorted example
Availability heuristic – Over-weigh easily available information in making decisions
Backfire effect – React to disconfirming evidence by strengthening beliefs
Belief bias – Evaluation of logical strength of argument based on belief in conclusion
Bias blind spot – I'm less biased than others
Choice-supportive bias – Remembering one's choices as better than they were
Superiority bias – Overestimate one's positive qualities relative to those of others
Hindsight bias – I knew it all along
Publication bias – Positive results more likely to be published than negative ones

If you think you are unbiased, you are

either a SAINT

or an ALIEN



Bad things happen only when bad people are involved.

This viewpoint is in contradiction with

- Accepted understanding in modern psychology and sociology
- The history of our species

Biases are ubiquitous, well-studied and documented.

Bad things happen with the best people and the best intentions.

Some issues

Off with its head! -- Reviewers like to REJECT, not accept
Un-falsifiable reviews -- Nice reviewers cannot dislodge mean ones
The clique effect -- PC members prefer papers by friends
Normal science – Rejection of critiques and alternative/novel viewpoints
Rule by consensus – Incremental preferred over ground-breaking
And more -- Reviews that are incorrect, incompetent, irresponsible …

Reviewers like to REJECT papers

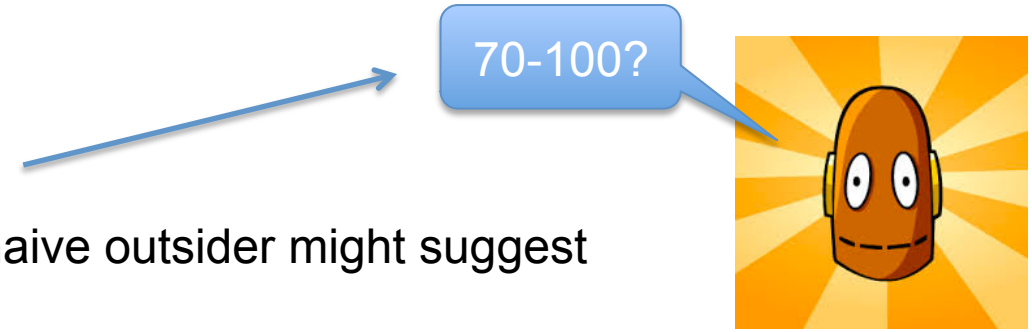
not ACCEPT papers

200 submissions.

Target 50 accepts, rate = 25%

1: Reject.
2: Lean towards reject
3: Undecided.
4: Lean towards accept
5: Accept. A solid paper.
6: Strong, exciting paper.

Q: How many submissions have average score ≥ 5 after 1st round of reviews?



What a naive outsider might suggest

Reviewers like to REJECT papers

not ACCEPT papers

200 submissions.

Target 50 accepts, rate = 25%

1: Reject.
2: Lean towards reject
3: Undecided.
4: Lean towards accept
5: Accept. A solid paper.
6: Strong, exciting paper.

Q: How many submissions have average score ≥ 5 after 1st round of reviews?

A: Typically 0-10

I think the paper is ok but I won't fight for it.
Fine paper but not above the bar for CRYPTO.

Some PC members do not give an accept score to any paper.

After the top 10% of papers, reviewers don't feel strongly about accepting anything. It is a crapshoot.

This is GOOD thing. It means we have HIGH STANDARDS.

I don't think that is what it means …

Reviewers like to REJECT papers

It is different in some other communities.

Reviewers like to REJECT papers

But $\text{Reviewers} \subset \text{Authors}$

Superiority bias

$$\text{Reviewers} \subset \text{Authors}$$

Most reviewers in our community think their own work is (much) better than that of their peers.

Superiority bias

Lake Wobegon effect: All the children are above average.

Most people think they are above-average drivers.

Our culture incentivizes and perpetuates rejection

- Negativity makes the reviewer seem smart
- No incentive to fight for a paper
- We review as we were reviewed

If you want the PC to think you are smart and well informed, be negative.

Mean reviewer

High standards, well informed, technically sophisticated.

Low standards, ignorant, technically weak.

Nice reviewer

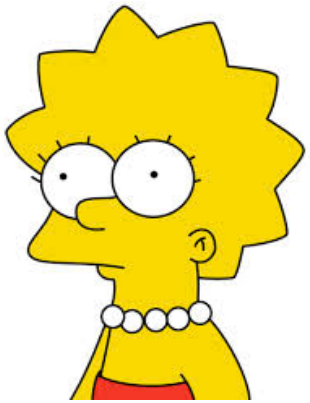
Weak paper. Minor, un-interesting results, low novelty. Incremental techniques.

Good paper. Strong, interesting, novel results. Novel techniques.

Our culture incentivizes and perpetuates rejection

- Negativity makes the reviewer seem smart
- No incentive to fight for a paper
- We review as we were reviewed

I don't want to antagonize mean reviewer. Other reviewers know my identity. The authors do not. So fighting for the paper can hurt me but agreeing to reject costs me nothing.



Nice, smart, but young reviewer

Our culture incentivizes and perpetuates rejection

- Negativity makes the reviewer seem smart
- No incentive to fight for a paper
- We review as we were reviewed

Reviews he got on his last four submissions

When I am a reviewer I must be very critical. Clearly a reviewer's job is to find reasons to reject.

REJECT: Proofs are in appendices, there is no Conclusions section, fails to cite ePrint paper, is un-interesting, un-surprising, has already been extended, …

What does all this even mean?
And what does it have to do with quality?

Not surprising

Not interesting

Trivial

If you want surprises …

Not surprising

Not interesting

Trivial

These reviewer comments are

Not falsifiable

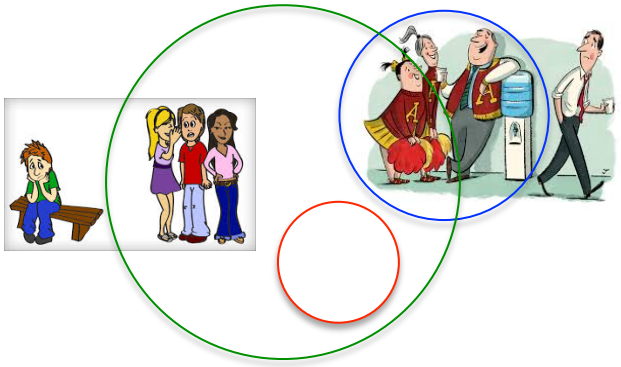
Intimidating

Clique

/klēk,klik/

noun

Small group of people with a common culture and shared interests who work together.



Our community is a collection of intersecting cliques. Clique size can be as small as 5. Often centered on a current topic.

PhD from same place
Advisor-student relations
History, Friendship

→ Clique

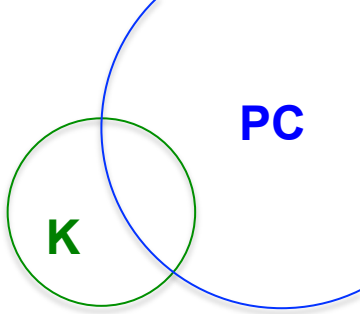
The Clique Effect

Many people in our community believe the clique effect is real and happens.

It can be observed.

Clique **K** well represented on PC

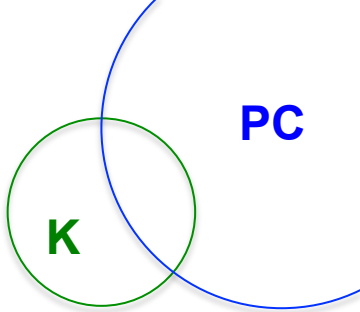
Papers by members of **K** will be more likely to be accepted than papers of non-members.



Explaining the Clique Effect

Clique **K** well represented on PC

Papers by members of **K** will be more likely to be accepted than papers of non-members.



The clique effect is not due to a conspiracy amongst clique members. It happens automatically due to the common culture, shared background and shared values of clique members.

C. Wright Mills
1916-1962
Sociologist

The Clique Effect

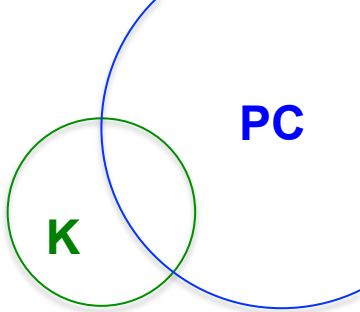
Game BAD

Clique members conspire and collude, de-anonymize their papers to each other, and accept mostly papers by members of the clique.

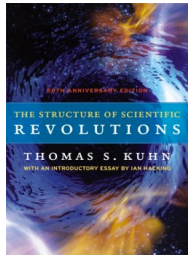
Game GOOD

Clique members evaluate papers independently, rationally and fairly from their perspective and select the ones that have the most scientific merit from their perspective.

These two games have indistinguishable outcomes



Normal science



Thomas Kuhn

Research firmly based on one or more prior scientific achievements acknowledged as providing foundations.

Students are mentored by researchers in these foundations. Seldom any disagreement over fundamentals.

Researchers investigate the kinds of questions to which their theories can provide answers.

Research turns into puzzle solving.

Opposition to, and rejection of, critiques and novel viewpoints.

Rule by consensus

Decisions taken largely based on consensus and average score

Accepted papers are the ones nobody hates rather than ones someone likes.

Incremental, mediocre work will dominate on the borderline.

But papers that have character often critique or challenge, and thus offend someone …

If a paper doesn't offend SOMEONE it can't have real character …

How should I review?

What is the ideal model functionality?

How should I review?

How should I live?

The golden rule of life:

Treat others as you would yourself wish to be treated

How should I review?

How should I live?

The golden rule of reviewing:

Review the papers of others as you would wish your own to be reviewed

Succinct guidelines for Reviewers

The preceding may (or may not) help to explain and understand some phenomena in the reviewing culture, but this is unlikely to change anything

because, even if most of us agree that bad reviews exist, few of us think of ourselves as ever providing one.

Bias blind spot

The fundamental problem with the reviewing system …

No Accountability

No place to appeal a decision
No way to overturn a decision
No consequences for reviewer actions

The President of the USA can be impeached.
There is nothing one can do to PC members.

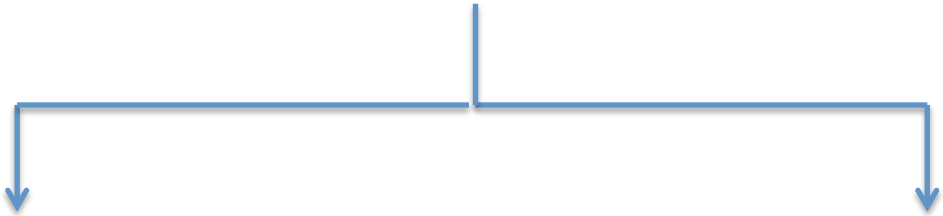
History has shown that power must be balanced by accountability to prevent abuse.

Peer review is a broken, dark ages system

because

It is fundamentally at odds with human nature and history.

Processes for decision making and judgment



Real world

Weak, realistic assumptions:
People are biased

↓

Checks and balances
Accountability

Our world

Strong, unrealistic assumptions:
People (reviewers) are unbiased

↓

Trust in authorities
No accountability

Works better in practice

Only works in theory

Peer review is a judicial system

A Model for Peer Review:

A court of law

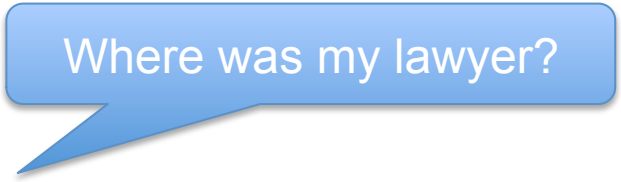
Court of law	Conference peer review
The accused	The submission
Decision = guilty, not guilty	Decision = reject, accept
Panel of judges	Program Committee (PC)
Chief Justice	PC Chair
Witnesses	Sub-reviewers

Court of law	Conference peer review
The accused	The submission
Decision = guilty, not guilty	Decision = reject, accept
Panel of judges	Program Committee (PC)
Chief Justice	PC Chair
Witnesses	Sub-reviewers
Advocate for defense	**[None]**
Public debates and opinions	**Secret debates and opinions**
Public review of judge appointments	**No public review of judge appointments**
Judge appointments by external parties	**Chief justice appoints the rest of the panel**
Decisions can be appealed	**No appeal for decisions**

Court of law	Conference peer review
The accused	The submission
Decision = guilty, not guilty	Decision = reject, accept
Panel of judges	Program Committee (PC)
Chief Justice	PC Chair
Witnesses	Sub-reviewers
Advocate for defense	**rebuttal?**
Public debates and opinions	**Secret debates and opinions**
Public review of judge appointments	**No public review of judge appointments**
Judge appointments by external parties	**Chief justice appoints the rest of the panel**
Decisions can be appealed	**Re-submission?**

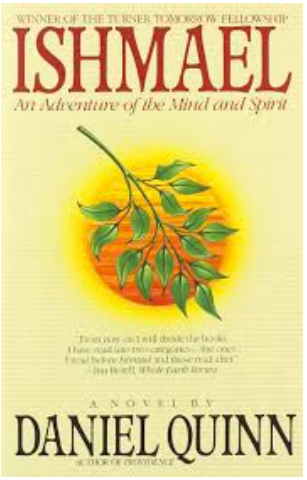
Do we want LAWYERS in research?

Does not cite ePrint paper – But it appeared after submission deadline
Is implied by submission 211 – But 211 is follow-up
Does not explain notation – See page 4
Is wrong – no it isn't
It is known – show me the reference



What's the solution?

Treat it as a research problem.
Think, write, talk, experiment.



Our community is creative and imaginative.
We have never shied away from hard problems. We have solved many.
This is another.

Create experimental publication venues.
Try out new reviewing systems.

Look elsewhere for ideas:

- Olympics: Highest and lowest scores are discarded
- Kahneman: Automation + narrow reviewer input
- …

Our disciplinary culture

is important and intriguing

We benefit from making it an explicit object of study and research.

Disciplines external to ours have much to offer.