Kernel-Size Lower Bounds: The Evidence from Complexity Theory

Andrew Drucker

IAS

Worker 2013, Warsaw

Part 3/3

Note

These slides are taken (with minor revisions) from a 3-part tutorial given at the 2013 Workshop on Kernelization ("Worker") at the University of Warsaw. Thanks to the organizers for the opportunity to present!

Preparation of this teaching material was supported by the National Science Foundation under agreements Princeton University Prime Award No. CCF-0832797 and Sub-contract No. 00001583. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Background

- Recall: [Fortnow-Santhanam '08] gave strong evidence for the OR-conjecture (for deterministic reductions).
- Left open:
  1. bounding power of two-sided bounded-error compressions of $\mathrm{OR}_{=}(L)$;
  2. any strong evidence for the AND-conjecture.
- Recently, success on both items. ([D. '12], this talk)

To be proved

Theorem (D.'12, special case)

Assume $\mathrm{NP} \not\subseteq \mathrm{coNP/poly}$. *If* L *is* NP*-complete and* $t(n) \leq \mathrm{poly}(n)$,

then no PPT reduction R *from either of*

$$\mathrm{OR}_{=}(L)^{t(\cdot)} \ , \ \mathrm{AND}_{=}(L)^{t(\cdot)}$$

to any problem L', *with* $\Pr[\text{success}] \geq .99$, *can achieve*

$$|R(\overline{x})| \ \leq t(n) \ .$$

To be proved

Theorem (D.'12, special case)

Assume $\mathrm{NP} \not\subseteq \mathrm{coNP/poly}$. *If* L *is* NP*-complete and* $t(n) \leq \mathrm{poly}(n)$,

then no PPT reduction R *from*

$$\mathrm{AND}_{=}(L)^{t(\cdot)}$$

to <u>*any problem*</u> L', *with* $\Pr[\text{success}] \geq .99$, *can achieve*

$$|R(\overline{x})| \leq .01t(n) \ .$$

Our goal

- Assume such an R does exist.
  We'll describe how to use reduction R for $\mathrm{AND}_{=}(L)$ to prove membership in $\overline{L}$.

- Initial protocol idea will be an interactive proof system to witness $x \in \overline{L}$.

- This can be converted to an NP/poly protocol for $\overline{L}$ by standard results.

  Thus $L \in \mathrm{coNP/poly}$; and L is NP-complete.

First, a story to motivate our approach. A story about... apples.[1]

[1] In the tutorial I just told the story out loud. It might seem a little silly put right on the slides; but I think it has pedagogical value.

Some apples taste good, some taste bad.

But you're allergic to apples.

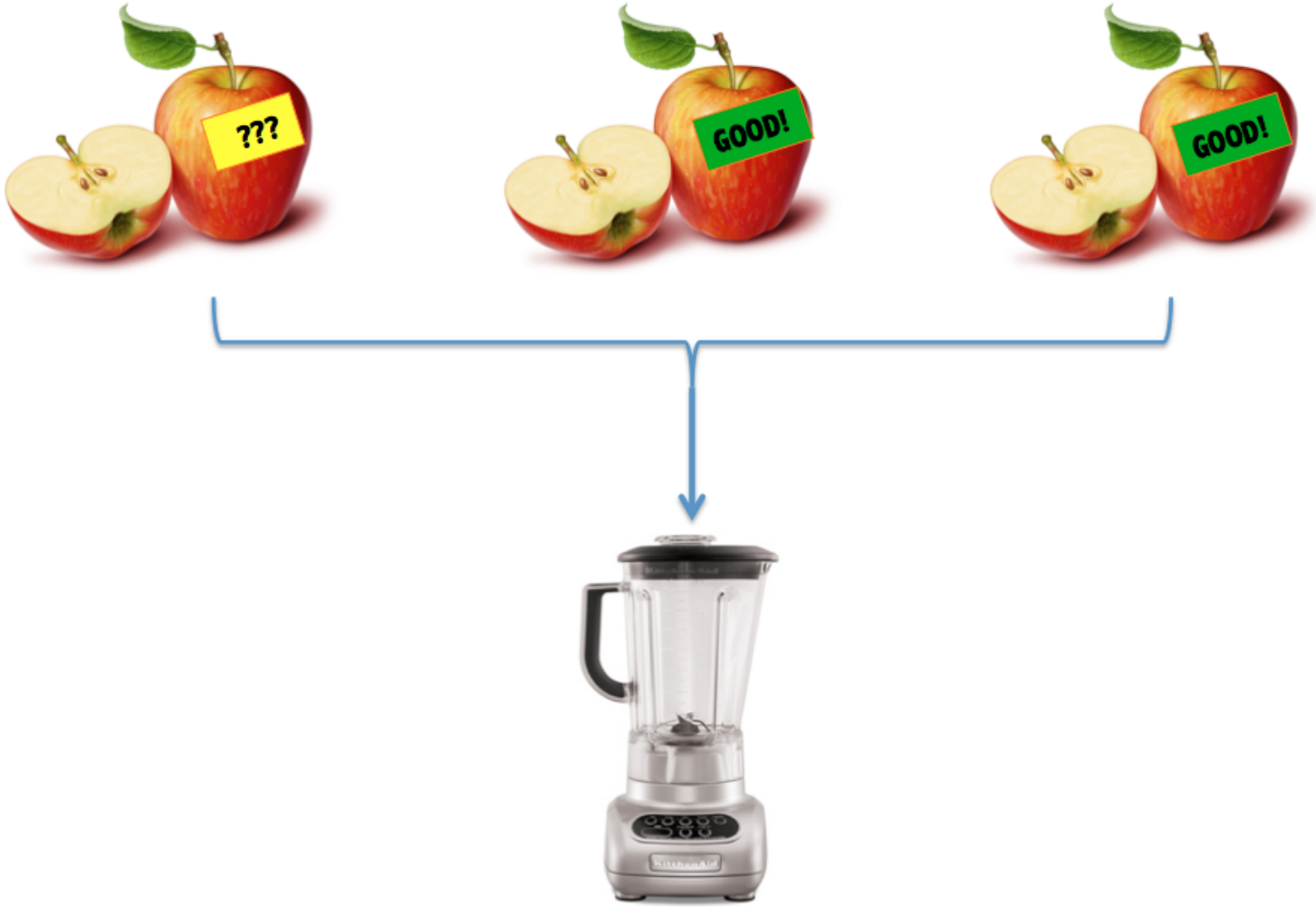
You can't eat them, so you can't tell good from bad directly.

That's where Merlin comes in.

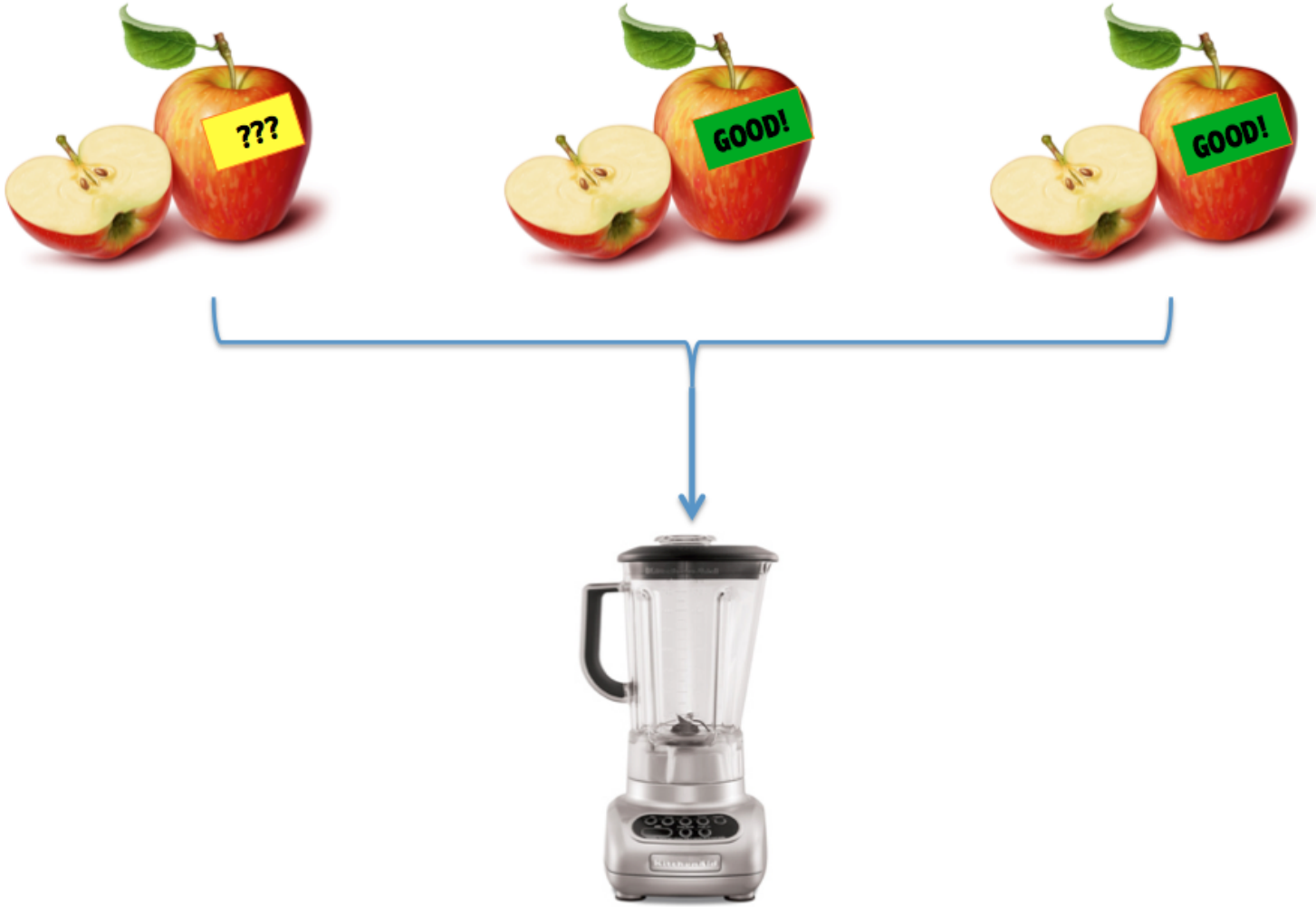
Merlin has a particular apple he really wants to convince you is bad.

But you don't trust Merlin. So what do you do?

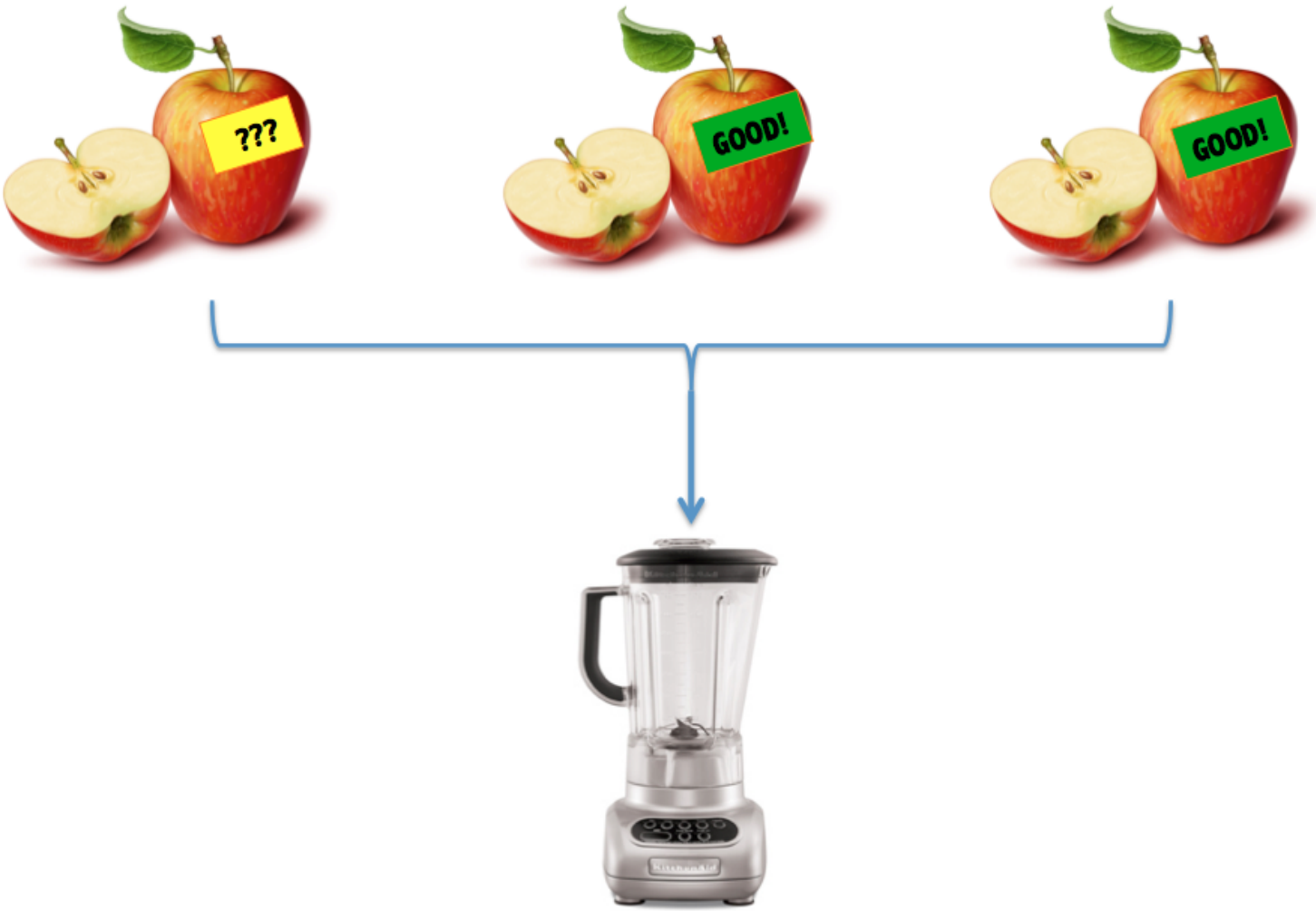
First, you get a blender.



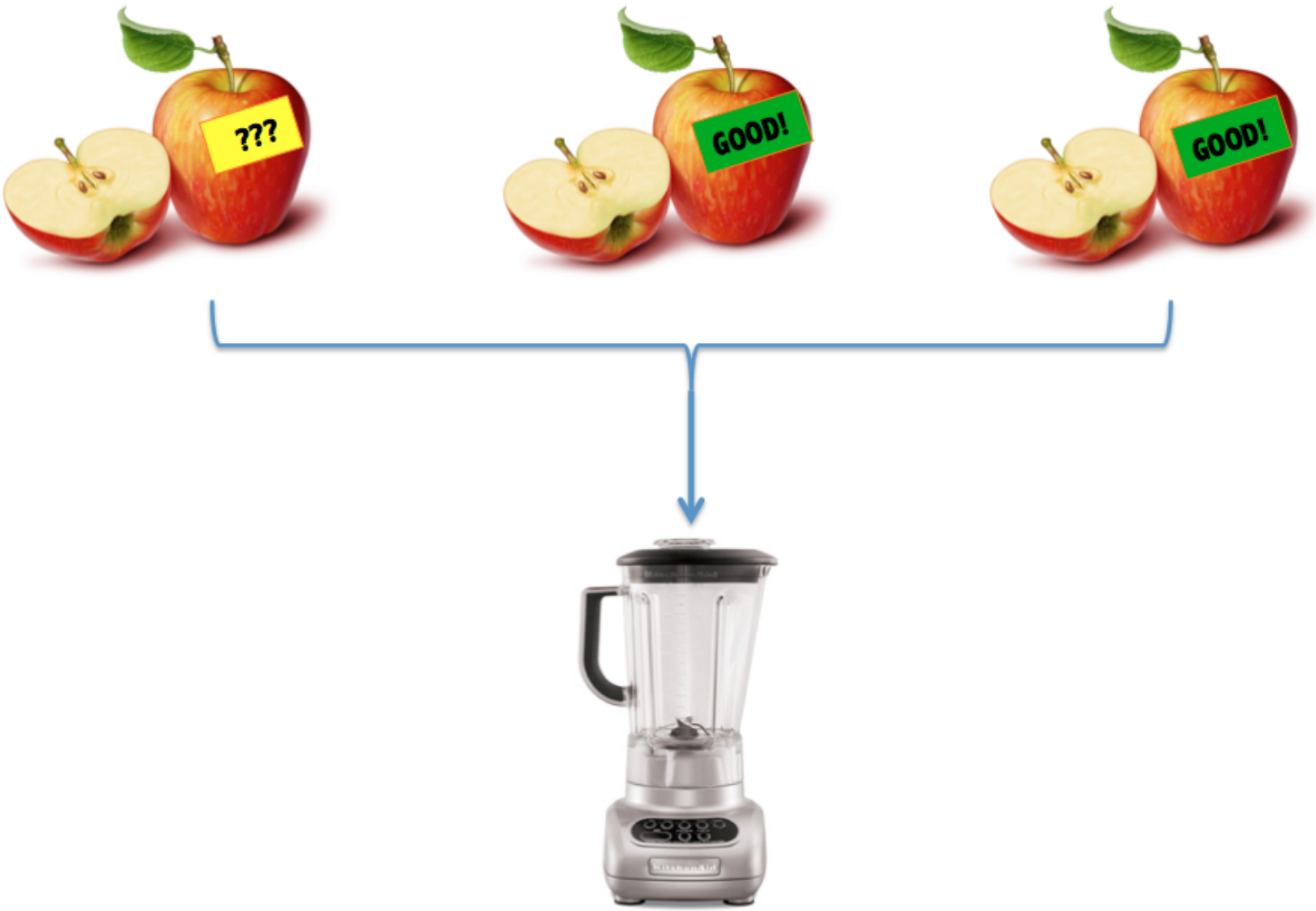
You throw Merlin's apple into a blender with a bunch of other apples, known to be good.



The result is a smoothie.
It will taste good exactly if all of the "input" apples are good.



You feed it to Merlin, and ask him if it tastes good.



But what will Merlin say, if he knows you used his apple?


Blecch!! That's awful!

So how do you make it harder for Merlin to lie?

You privately flip a coin.
Heads, you include Merlin's apple.
Tails, you include only known good apples.

If Merlin's apple really is bad,
he'll be able to taste whether we used it.

Now suppose Merlin is lying, and his apple is good.

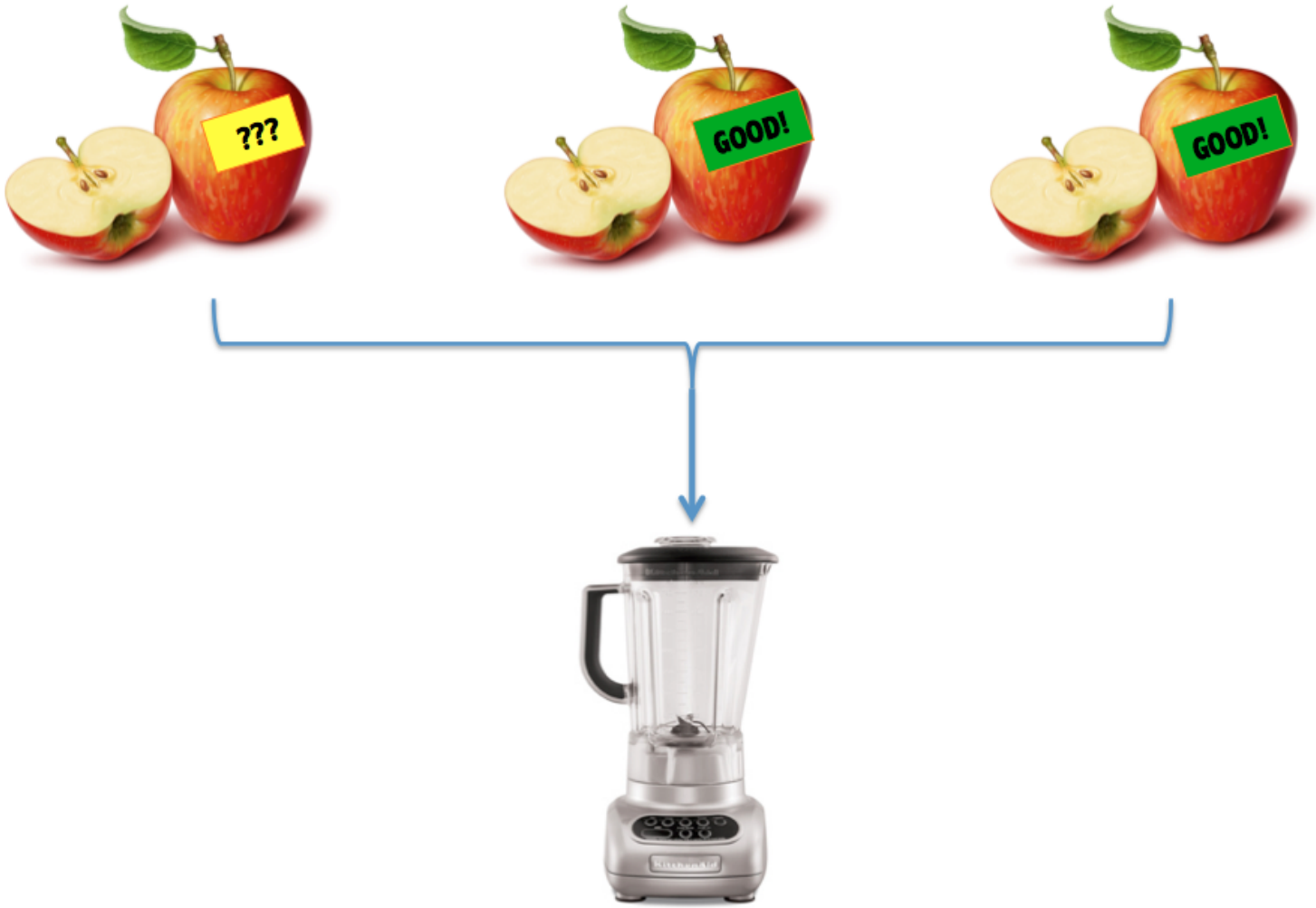
Then the smoothies taste good in
either case, and Merlin is confused!


Umm.. That
tastes... good?

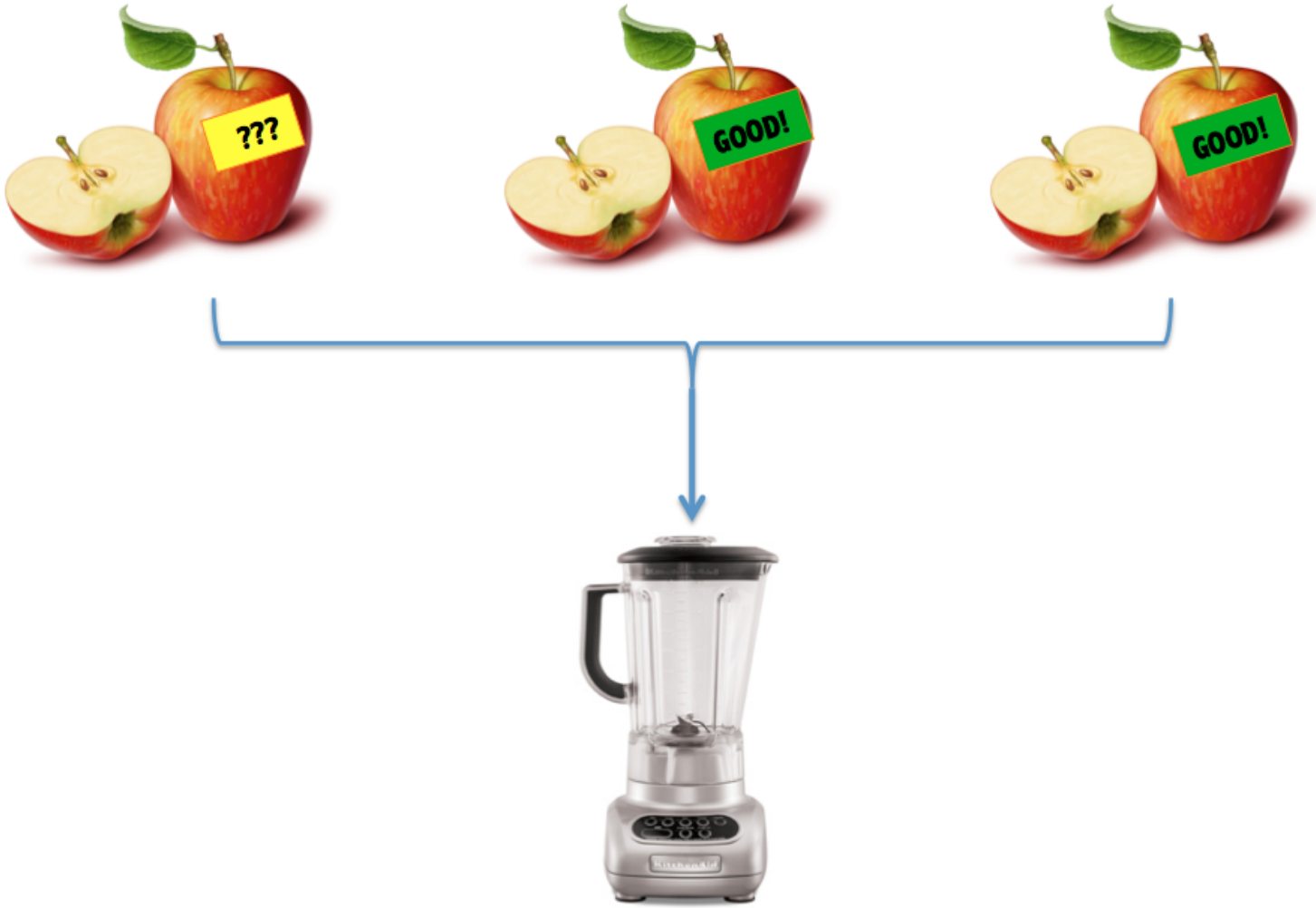
Can't reliably tell you if his apple was used.

But life is not quite so simple.

First, if the blender isn't powerful enough, it might leave chunks of Merlin's apple he can identify. Would help him to lie.



Second, if Merlin's apple is a Granny Smith,
and all your apples are Red Delicious,
he might again taste the difference (even if Merlin's apple is good).



Thus, you will need a sufficient diversity of good apples, and may also want to randomize which of your apples you throw in.

- All this is a metaphorical description of our basic strategy, by which we'll use a compression reduction for $\mathrm{AND}_{=}(L)$ to build an interactive proof system for $\overline{L}$.

- Apples correspond to inputs x to the decision problem for L. Merlin is trying to convince us that a particular x^* lies in $\overline{L}$.

- Apples' goodness corresponds to membership in L. Merlin claims the "apple" x^* is bad.

- The blender represents a compression reduction for $\mathrm{AND}_{=}(L)$. We will test Merlin's "distinguishing ability" just as described.

- A "powerful" blender, leaving few chunks, corresponds to a reduction achieving strong compression.

- The need for diverse "input" apples will correspond to a need to have diverse elements of L to insert into the compression reduction along with x^*.

- Hopefully this story will be helpful in motivating what follows.
- Now, we need to shift gears and develop some math background for our work.

Math background

- **Review:** minimax theorem; basic notions from probability, information theory.
- **Recall:** 2-player, simul-move, zero-sum games.

Theorem (Minimax)

Suppose in game $\mathbf{G} = (X, Y, \mathrm{Val})$*, for each P2 mixed strategy* $\mathcal{D}_Y$*, there is a P1 move* x *such that*

$$\mathbb{E}_{\mathbf{y} \sim \mathcal{D}_Y}[\ \mathrm{Val}(x, \mathbf{y})\ ] \ \leq \ \alpha\ .$$

Then, there is a P1 mixed strategy $\mathcal{D}^*_X$ *such that, for every P2 move* y*,*

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{D}^*_X}[\ \mathrm{Val}(\mathbf{x}, y)\ ] \ \leq \ \alpha\ .$$

Probability distributions

- **Statistical distance of (finite) distributions:**

$$||\mathcal{D} - \mathcal{D}'|| = \frac{1}{2}\sum_{u} |\mathcal{D}(u) - \mathcal{D}'(u)|$$

- Also write $||X - X'||$ for random variables.
- Alternate, "distinguishing characterization" often useful...

Probability distributions

Distinguishing game

Arthur: $b \in_r \{0,1\}$; *samples*

$$u \sim \begin{cases} \mathcal{D} & \text{if } b = 0, \\ \mathcal{D}' & \text{if } b = 1. \end{cases}$$

Merlin: *receives* u, *outputs guess for* b.

Claim

Merlin's maximum success prob. is

$$\mathrm{suc}^* = \frac{1}{2}\left(1 + ||\mathcal{D} - \mathcal{D}'||\right) .$$

Entropy and information

- **Entropy** of a random variable:

$$H(X) \;:=\; \sum_x \Pr[X = x] \cdot \log_2 \left(\frac{1}{\Pr[X = x]} \right)$$

  Measure of information content of X...

- Same def. works for joint random vars, e.g. $H(X, Y)$.

- **Mutual information** between random vars:

$$I(X; Y) \;:=\; H(X) + H(Y) - H(X, Y)\ .$$

  "how much X tells us about Y" (and vice versa)

Entropy and information

- **Mutual information** between random vars:

$$I(X;Y) \;:=\; H(X) + H(Y) - H(X,Y)\ .$$

- **Examples:**
  1. X, Y independent $\quad\Longrightarrow\quad I(X;Y) = 0$;
  2. $X = Y$ $\quad\Longrightarrow\quad I(X;Y) = H(X)$.

- Always have $0 \leq I(X;Y) \leq H(X), H(Y)$.

Entropy and information

Question: which is bigger,

$$I(X^1, X^2 \,;\, Y) \qquad \text{or} \qquad I(X^1; Y) + I(X^2; Y) \quad ?$$

(Consider cases...)

Entropy and information

Claim

Suppose $\overline{X} = X^1, \ldots, X^t$ *are* independent *r.v.'s. Then,*

$$I\left(\overline{X}; Y\right) \geq \sum_j I(X^j; Y) \ .$$

- **Intuition:** Information in X^i about Y is "disjoint" from info in X^j about Y...

Conditioning

- Let X, Y be jointly distributed r.v.'s.
- $X_{[Y=y]}$ denotes X conditioned on $[Y = y]$.
- $I(X;Y)$ small means conditioning has little effect:

Claim

For any X, Y,

$$\mathbb{E}_{x\sim X}||Y_{[X=x]} - Y|| \leq \sqrt{I(X;Y)}.$$

(follows from "Pinsker inequality")

Conditioning

Claim

For any X, Y,

$$\mathbb{E}_{x \sim X} || Y_{[X=x]} - Y || \leq \sqrt{I(X;Y)} .$$

Example [BBCR'10]: let $X^1, \ldots, X^t$ be uniform, and

$$Y = \mathrm{MAJ}(X^1, \ldots, X^t) .$$

Then:

1. $I(X^1; Y) \leq 1/t$;
2. $\left|\left| Y - Y_{[X^1=b]} \right|\right| \approx 1/\sqrt{t}$.

Key lemma

A fact about statistical behavior of compressive mappings:

Lemma (Distributional stability—binary version)

Let $F : \{0,1\}^t \to \{0,1\}^{t'<t}$ *be given. Let* $F(\mathcal{U}_t)$ *denote output dist'n on uniform inputs, and*

$$F(\mathcal{U}_t|_{j \leftarrow b})$$

denote output distribution with j^{th} *input fixed to* b*. Then,*

$$\mathbb{E}_{\mathbf{j} \in_r [t],\ b \in_r \{0,1\}} \ || \ F(\mathcal{U}_t|_{\mathbf{j} \leftarrow b}) \ - \ F(\mathcal{U}_t) \ || \ \leq \ \sqrt{t'/t} \ .$$

Proof.

Follows from previous two Claims (and Jensen ineq). □

Key lemma

A fact about statistical behavior of compressive mappings:

Lemma (Distributional stability—binary version)

Let $F : \{0,1\}^t \to \{0,1\}^{t' < t}$ *be given. Let* $F(\mathcal{U}_t)$ *denote output dist'n on uniform inputs, and*

$$F(\mathcal{U}_t|_{j \leftarrow b})$$

denote output distribution with j^{th} *input fixed to* b*. Then,*

$$\mathbb{E}_{\mathbf{j} \in_r [t],\ b \in_r \{0,1\}} \ || \ F(\mathcal{U}_t|_{\mathbf{j} \leftarrow b}) \ - \ F(\mathcal{U}_t) \ || \ \leq \ \sqrt{t'/t} \ .$$

Similar lemmas and proof used, e.g., in [Raz'95] on parallel repetition. R. Impagliazzo, A. Nayak, S. Vadhan helped me understand the proof going through mutual information and Pinsker ineq. My original proof in [D'12] used a different approach, based on encoding/decoding and Fano's inequality.

Back to business

- **Recall:** L is NP-complete, $t(n) \leq \text{poly}(n)$, and R reduces an AND of $t(n)$ L-instances to a short, equivalent L-instance, success prob. $= .99$.

- (again, assuming here that target problem $L' = L$)

Initial setting

- Fix attention to a single input size $n > 0$. Fix $t := t(n) \leq \text{poly}(n)$.
- The PPT reduction

$$R(\overline{x}) \;=\; R(x^1, \ldots, x^t) : \quad \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$$

satisfies: $\forall \overline{x}$,

$$\bigwedge_j [x^j \in L] \quad \Longrightarrow \Pr_R[\, R(\overline{x}) \;\in\; L \,] \;\geq\; .99 \;,$$

$$\exists\, x^j \in \overline{L} \quad \Longrightarrow \Pr_R[\, R(\overline{x}) \;\in\; L \,] \;\leq\; .01 \;.$$

Game plan

- **Basic observation:** suppose

$$x^1, \ldots, x^t \in L_n \ ,$$

$$x \in \overline{L}_n \ .$$

(color-coded!)

- Consider the two computations

$$R(x^1, \ldots, x^t) \ , \qquad R(x^1, \ldots, \underbrace{x}_{(\text{coord. } j)}, \ldots, x^t)$$

Game plan

- **Basic observation:** suppose

$$x^1, \ldots, x^t \in L_n \ ,$$

$$x \in \overline{L}_n \ .$$

(color-coded!)

- Consider the two computations

$$R(\overline{x}) \ , \qquad R(\ \overline{x}[x; j]\ )$$

(for brevity)

Game plan

- **Observation:** the output distributions

$$R(\overline{x}) \ , \qquad R(\ \overline{x}[x;j]\ )$$

are far apart in statistical distance!

- first usually in L, second usually in $\overline{L}$...

Game plan

- **"Boosted" observation:** for any distribution $\overline{\mathcal{D}}$ over L_n^t, the output distributions

$$R(\ \overline{\mathcal{D}}\ )\ , \qquad R(\ \overline{\mathcal{D}}[x; j]\ )$$

  are far apart!

- We have:

$$\left|\left|\ R(\ \overline{\mathcal{D}}\ )\ \ -\ \ R(\ \overline{\mathcal{D}}[x; j]\ )\ \right|\right|\ \geq\ .98\ .$$

Game plan

- **Plan:** let $x \in \{0,1\}^n$ be a string; we wish to be convinced that $x \notin L$.
- The distributions

$$R(\,\overline{\mathcal{D}}\,)\,, \qquad R(\,\overline{\mathcal{D}}[x;j]\,)$$

  are far apart; but may computationally hard to distinguish.

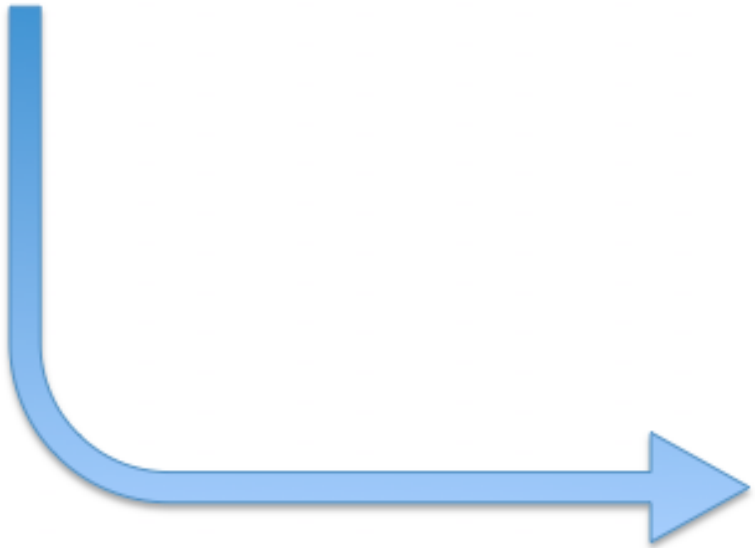
- **So:** we will ask Merlin to distinguish them!

A distinguishing task

$R(\overline{\mathcal{D}})$ $\quad$ $R(\overline{\mathcal{D}}[x; j])$

A distinguishing task

$R(\overline{\mathcal{D}})$ $R(\overline{\mathcal{D}}[x;j])$

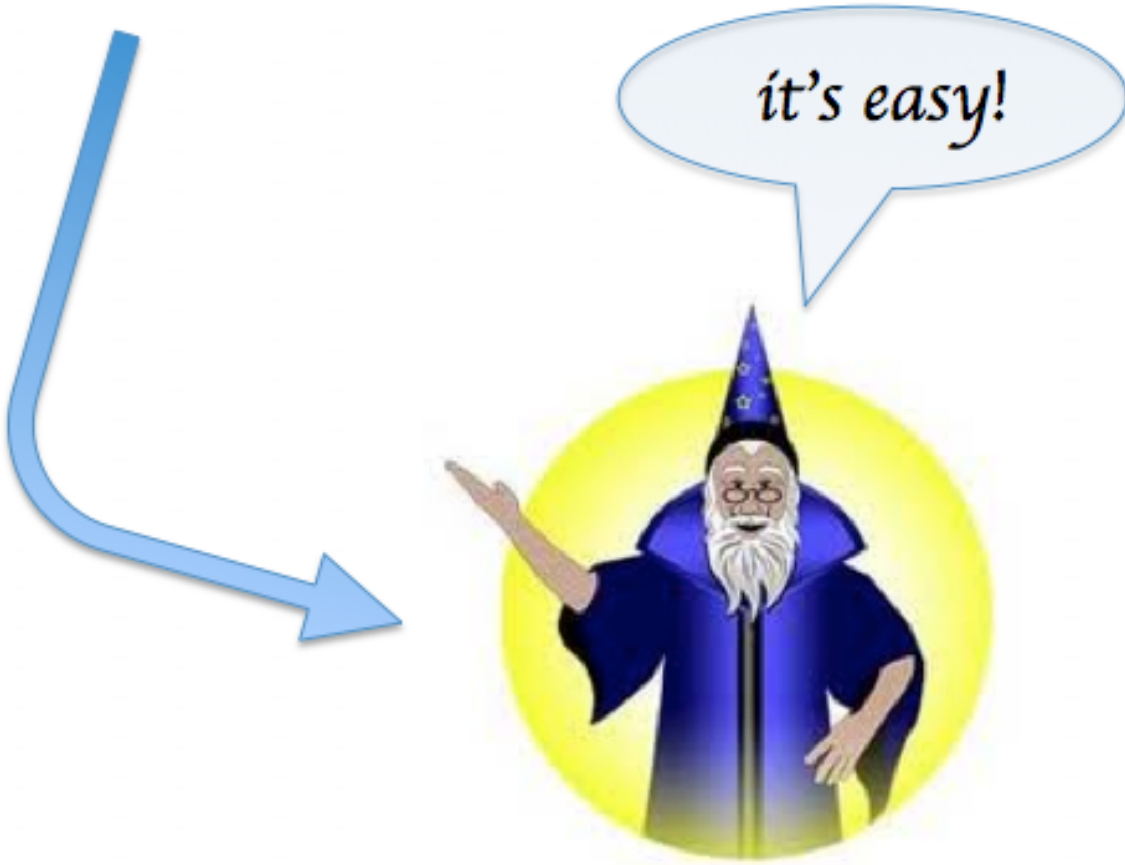


A distinguishing task

$R(\overline{\mathcal{D}})$ $R(\overline{\mathcal{D}}[x;j])$

A distinguishing task

$R(\overline{\mathcal{D}})$ $R(\overline{\mathcal{D}}[x;j])$



A distinguishing task

$R(\overline{\mathcal{D}})$ $\qquad$ $R(\overline{\mathcal{D}}[x; j])$

- **Main question:** how to choose $\overline{\mathcal{D}}$ and j?
- **Want:** for all $x \in L$, Merlin should be unable to distinguish between

$$R(\overline{\mathcal{D}}), \qquad R(\overline{\mathcal{D}}[x; j])$$

A distinguishing task

- **Also want:** $\overline{\mathcal{D}}$ sampleable efficiently using $\mathrm{poly}(n)$ advice.
- **Main technical lemma:** Such a $\overline{\mathcal{D}}$ can be constructed!

The main lemma

Lemma ("Disguising Distributions")

Given any mapping $R : \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$ *and language* L, *there exists a distribution* $\overline{\mathcal{D}^*}$ *over* L_n^t *such that:*

- *for any* $x \in L_n$,
- *if* $\mathbf{j} \in_r [t]$ *is uniformly chosen,*

$$\mathbb{E}_{\mathbf{j}} \left|\left| R(\overline{\mathcal{D}}^*) - R(\overline{\mathcal{D}}^*[x;\mathbf{j}]) \right|\right| \leq .3$$

Moreover, $\overline{\mathcal{D}^*}$ *can be sampled by a* $\mathrm{poly}(n)$*-sized circuit.*

The main lemma

Lemma ("Disguising Distributions"—general)

Given any mapping $R : \{0,1\}^{n \times t} \to \{0,1\}^{t'}$ *and language* L*, there exists a distribution* $\overline{\mathcal{D}^*}$ *over* L_n^t *such that:*

- *for any* $x \in L_n$*,*
- *if* $\mathbf{j} \in_r [t]$ *is uniformly chosen,*

$$\mathbb{E}_{\mathbf{j}} \left|\left| R(\overline{\mathcal{D}}^*) - R(\overline{\mathcal{D}}^*[x;\mathbf{j}]) \right|\right| \leq O(\sqrt{t'/t})$$

Moreover, $\overline{\mathcal{D}^*}$ *can be sampled by a* $\mathrm{poly}(n)$*-sized circuit.*

The main lemma

Lemma ("Disguising Distributions"—general, alternative bound)

Given any mapping $R : \{0,1\}^{n \times t} \to \{0,1\}^{t'}$ *and language* L*, there exists a distribution* $\overline{\mathcal{D}^*}$ *over* L_n^t *such that:*

- *for any* $x \in L_n$*,*
- *if* $\mathbf{j} \in_r [t]$ *is uniformly chosen,*

$$\mathbb{E}_{\mathbf{j}} \left\| R(\overline{\mathcal{D}}^*) - R(\overline{\mathcal{D}}^*[x;\mathbf{j}]) \right\| \leq 1 - 2^{-O(t'/t)}$$

Moreover, $\overline{\mathcal{D}^*}$ *can be sampled by a* $\mathrm{poly}(n)$*-sized circuit.*

The main lemma

Lemma ("Disguising Distributions")

Given any mapping $R : \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$ *and language* L, *there exists a distribution* $\overline{\mathcal{D}^*}$ *over* L_n^t *such that:*

- *for any* $x \in L_n$,
- *if* $\mathbf{j} \in_r [t]$ *is uniformly chosen,*

$$\mathbb{E}_{\mathbf{j}} \left\| R(\overline{\mathcal{D}}^*) - R(\overline{\mathcal{D}}^*[x;\mathbf{j}]) \right\| \leq .3$$

Moreover, $\overline{\mathcal{D}^*}$ *can be sampled by a* $\mathrm{poly}(n)$*-sized circuit.*

- **Intuition:** $x \in L_n$ is being tossed into R with "others like it" (all in L_n)...
- R highly compressive, so "forgets" most of its input.
  $\to$ We'll force it to forget about x!

Building disguising distributions

- Say that dist'n $\overline{\mathcal{D}}$ over L_n^t **disguises** a string $x \in L_n$ if

$$\mathbb{E}_{\mathbf{j}} \left|\left| R(\,\overline{\mathcal{D}}^*\,) \;-\; R(\,\overline{\mathcal{D}}^*[x;\mathbf{j}]\,) \right|\right| \;\leq\; .3$$

- Need to find samplable $\overline{\mathcal{D}^*}$ that disguises all x.

- Seems hard... hope to apply minimax theorem to make things easier!

Building disguising distributions

Define (another) 2-player, simul-move game between

P1 ("Maker") and P2 ("Breaker"). Fix a large $M \leq \text{poly}(n)$.

Game

- **P1:** *Chooses a dist'n* $\overline{\mathcal{D}}$ *over* L_n^t *sampled by a ckt of size* M.
- **P2:** *Chooses an* $x \in L_n$.
- **Payoff to P2:**

$$\alpha \;:=\; \mathbb{E}_{\mathbf{j}} \;\big|\big|\; R(\,\overline{\mathcal{D}}\,) \;\;-\;\; R(\,\overline{\mathcal{D}}[x;\mathbf{j}]\,) \;\big|\big|$$

(Potential for confusion: P1's pure strategies are distributions...)

Building disguising distributions

Game

- **P1:** *Chooses a dist'n* $\overline{\mathcal{D}}$ *over* L_n^t *sampled by a ckt of size* M.
- **P2:** *Chooses an* $x \in L_n$.
- **Payoff to P2:**

$$\alpha \;:=\; \mathbb{E}_{\mathbf{j}} \;\Big|\Big|\; R(\;\overline{\mathcal{D}}\;) \;\;-\;\; R(\;\overline{\mathcal{D}}[x;\mathbf{j}]\;) \;\Big|\Big|$$

- We'll show that **for every** P2 mixed strategy $\mathbf{x} \sim \mathcal{X}$, there exists a P1 move $\overline{\mathcal{D}}$ that causes $\mathbb{E}_{\mathbf{x}}[\alpha] \;\leq\; .25$.
- Then, minimax thm. implies: $\exists$ a dist'n $\overline{\mathfrak{D}}$ over dist'ns such that for all x,

$$\mathbb{E}_{\mathbf{j},\overline{\mathcal{D}}\sim\overline{\mathfrak{D}}} \;\Big|\Big|\; R(\;\overline{\mathcal{D}}\;) \;\;-\;\; R(\;\overline{\mathcal{D}}[x;\mathbf{j}]\;) \;\Big|\Big| \;\leq\; .25$$

Building disguising distributions

- $\exists$ a dist'n $\overline{\mathfrak{D}}$ over dist'ns such that for all x,

$$\mathbb{E}_{\mathbf{j},\overline{\mathcal{D}}\sim\overline{\mathfrak{D}}} \left|\left| R(\overline{\mathcal{D}}) - R(\overline{\mathcal{D}}[x;\mathbf{j}]) \right|\right| \leq .25$$

$\Longrightarrow$ for all x,

$$\mathbb{E}_{\mathbf{j}} \left|\left| R(\overline{\mathfrak{D}}) - R(\overline{\mathfrak{D}}[x;\mathbf{j}]) \right|\right| \leq .25$$

The fact we used:

Claim

Let $\{\mathcal{R}_v\}_v$, $\{\mathcal{R}'_v\}_v$

be two families of dist'ns, $\mathbf{v}$ *a random variable, and let* $\mathfrak{R}, \mathfrak{R}'$ *be obtained by sampling from* $\mathcal{R}_{\mathbf{v}}, \mathcal{R}'_{\mathbf{v}}$ *respectively. Then,*

$$||\mathfrak{R} - \mathfrak{R}'|| \leq \sum_v \Pr[\mathbf{v} = v] \cdot ||\mathcal{R}_v - \mathcal{R}'_v|| .$$

Building disguising distributions

- Minimax gave a P1 mixed strategy $\overline{\mathfrak{D}}$ such that, for all x,

$$\mathbb{E}_{\mathbf{j}} \left|\left| R(\,\overline{\mathfrak{D}}\,) \;-\; R(\,\overline{\mathfrak{D}}[x;\mathbf{j}]\,) \right|\right| \;\leq\; .25$$

- This $\overline{\mathfrak{D}}$ may not itself be sampleable in size M!

- But, forming a mixture of $O(n)$ samples from $\overline{\mathfrak{D}}$ yields a $\overline{\mathcal{D}}^*$ that is nearly as good, and of complexity $O(Mn) \leq \text{poly}(n)$.

  ("strategy-sparsification" concept:
  [Lipton-Young '94, Althofer '94])

What we need now

So: to build disguising distributions for R, we just need to prove:

Claim

For every dist'n $\mathcal{X}$ over L_n, $\exists$ a dist'n $\overline{\mathcal{D}}$ over L_n^t such that:

$$\mathbb{E}_{\mathbf{j},\mathbf{x}\sim\mathcal{X}} \left|\left| R(\,\overline{\mathcal{D}}\,) \;-\; R(\,\overline{\mathcal{D}}[\mathbf{x};\mathbf{j}]\,) \right|\right| \;\leq\; .25\;.$$

Will use simplification ideas of Holger Dell (pers. comm.)

Key lemma

Lemma (Distributional stability—binary version)

Let $F : \{0,1\}^t \to \{0,1\}^{t' < t}$ *be given. Let* $F(\mathcal{U}_t)$ *denote output dist'n on uniform inputs, and*

$$F(\mathcal{U}_t|_{j \leftarrow b})$$

denote output distribution with j^{th} *input fixed to* b*. Then,*

$$\mathbb{E}_{\mathbf{j} \in_r [t],\ b \in_r \{0,1\}} \ || \ F(\mathcal{U}_t|_{\mathbf{j} \leftarrow b}) \ - \ F(\mathcal{U}_t) \ || \ \leq \ \sqrt{t'/t} \ .$$

Using distributional stability

Corollary

Let $\mathcal{X}$ be over L_n. Let $x^1, \ldots, x^t, y^1, \ldots, y^t$ be $2t$ independent samples from $\mathcal{X}$, and let $\overline{\mathcal{D}}$ be uniform dist'n on

$$\{x^1, y^1\} \times \ldots \times \{x^t, y^t\} \quad \subset \quad L_n^t \; .$$

Then,

$$\mathbb{E}_{\mathbf{j} \in_r [t]} \; \left|\left| R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow x^{\mathbf{j}}}) - R(\overline{\mathcal{D}}) \right|\right| \; \leq \; \sqrt{.01} \; = \; .1 \; .$$

Proof: after fixing **any** tuples $\overline{x}, \overline{y}$, use Dist. Stability Lemma on induced function $F_{\overline{x},\overline{y}}$. Here $t' = .01t$.

Using distributional stability

Corollary

Let $\mathcal{X}$ be over L_n. Let $x^1, \ldots, x^t, y^1, \ldots, y^t$ be $2t$ independent samples from $\mathcal{X}$, and let $\overline{\mathcal{D}}$ be uniform dist'n on

$$\{x^1, y^1\} \times \ldots \times \{x^t, y^t\} \quad \subset \quad L_n^t \; .$$

Then,

$$\mathbb{E}_{\mathbf{j} \in_r [t]} \; \left|\left| R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow x^{\mathbf{j}}}) - R(\overline{\mathcal{D}}) \right|\right| \; \leq \; .1 \; .$$

Claim: w.h.p. the $\overline{\mathcal{D}}$ built above works as required P1 strategy, in response to P2 mixed strategy $\mathcal{X}$.

Idea: w.h.p. over construction, $\mathbf{x} \sim \mathcal{X}$, and $\mathbf{j}$, dist'ns

$$R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow x^{\mathbf{j}}}) \; , \quad R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow y^{\mathbf{j}}}) \; , \quad R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow \mathbf{x}})$$

are all close to $R(\overline{\mathcal{D}})$...

Using distributional stability

Corollary

Let $\mathcal{X}$ be over L_n. Let $x^1, \ldots, x^t, y^1, \ldots, y^t$ be $2t$ independent samples from $\mathcal{X}$, and let $\overline{\mathcal{D}}$ be uniform dist'n on

$$\{x^1, y^1\} \times \ldots \times \{x^t, y^t\} \quad \subset \quad L_n^t \, .$$

Then,

$$\mathbb{E}_{\mathbf{j} \in_r [t]} \; \left|\left| R(\overline{\mathcal{D}}|_{\mathbf{j} \leftarrow x^{\mathbf{j}}}) - R(\overline{\mathcal{D}}) \right|\right| \; \leq \; .1 \, .$$

Notice: to build an input-distribution $\overline{\mathcal{D}}$ to disguise the insertion of $\mathbf{x} \sim \mathcal{X}$, we used inputs that were "as similar to $\mathbf{x}$ as possible"—because drawn from the same distribution $\mathcal{X}$.

Makes sense as a strategy!

The upshot

- **Recall:** n, t are fixed and $R : \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$.
- We have used (minimax + sparsification) to produce a samplable dist'n $\overline{D}^*$ over L_n^t, such that for all $x \in L_n$,

$$\mathbb{E}_{\mathbf{j}} \, ||R(\, \overline{D}^* \,) \; - \; R(\, \overline{D}^*[x; \mathbf{j}] \,)|| \; \leq \; .3 \; .$$

- On the other hand, AND-property of R gives: for all $x \in \overline{L}_n$,

$$\mathbb{E}_{\mathbf{j}} \, ||R(\, \overline{D}^* \,) \; - \; R(\, \overline{D}^*[x; \mathbf{j}] \,)|| \; \geq \; .98 \; .$$

Now "hide the value of $\mathbf{j}$"... doesn't increase statistical distance in 1^{st} case, or affect argument in 2^{nd} case!

The upshot

- **Recall:** n, t are fixed and $R : \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$.
- We have used (minimax + sparsification) to produce a samplable dist'n $\overline{D}^*$ over L_n^t, such that for all $x \in L_n$,

$$||R(\ \overline{D}^*\ )\ -\ R(\ \overline{D}^*[x;\mathbf{j}]\ )||\ \leq\ .3\ .$$

- On the other hand, AND-property of R gives: for all $x \in \overline{L}_n$,

$$||R(\ \overline{D}^*\ )\ -\ R(\ \overline{D}^*[x;\mathbf{j}]\ )||\ \geq\ .98\ .$$

Now "hide the value of $\mathbf{j}$"... doesn't increase statistical distance in 1^{st} case, or affect argument in 2^{nd} case!

The upshot

What we've done so far: we built a reduction Q computable by $\text{poly}(n)$-sized circuits:

- **Input:** $x \in \{0,1\}^n$.
- **Output:** a pair of sampling-circuit descriptions

$$\langle C \ , \ C'_x \rangle$$

  where:
  - C samples from $R(\overline{D}^*)$,
  - C' samples from $R(\overline{D}^*[x;\mathbf{j}])$, $\quad \mathbf{j} \in_r [t]$.
- **Property:** if $x \in \overline{L}_n$, then

$$||C - C'_x|| \ \geq \ .98 \ ,$$

  while if $x \in L_n$,

$$||C - C'_x|| \ \leq \ .3 \ .$$

The upshot

- This, combined with the **Arthur/Merlin distinguishing protocol** mentioned earlier gives a (non-uniform)

  2-message, private-coin interactive proof system

  to witness membership in $\overline{L}$.

- By standard techniques [Goldwasser-Sipser '86, Babai '85, Adleman'78], this implies $\overline{L} \in \mathrm{NP/poly}$, i.e., $L \in \mathrm{coNP/poly}$.

- As L was NP-complete, we get $\mathrm{NP} \subset \mathrm{coNP/poly}$. Mission accomplished! So in fact the reduction R is unlikely to exist.

The statistical distance problem $\mathrm{SD}^{\geq .9}_{\leq .3}$

Problem (SD)

- **Input:** *sampling-circuits* $\langle C, C' \rangle$.
- **Distinguish:** *Case (i):* $||C - C'|| \geq .9$;
  Case (ii): $||C - C'|| \leq .3$.

This promise problem has 2-message interactive proof systems to prove we are in Case (i)—as mentioned. (Proof-of-distance)

$R(\ \overline{\mathcal{D}}\ )$ $R(\ \overline{\mathcal{D}}[x; j]\ )$

The statistical distance problem $\mathrm{SD}^{\geq .9}_{\leq .3}$

Problem (SD)

- **Input:** *sampling-circuits* $\langle C, C' \rangle$.
- **Distinguish:** *Case (i):* $||C - C'|| \geq .9$;
  Case (ii): $||C - C'|| \leq .3$.

- But, in fact, also has 2-message Proof-of-closeness interactive proof systems to prove we are in Case (ii)!
- Follows from results of [Fortnow '87], [Sahai-Vadhan '99] on zero-knowledge proofs.
- **This** $\Rightarrow$ hardness of probabilistic compression for $\mathrm{OR}_{=}(L)$...

Compression for $\mathrm{OR}_{=}(L)$

- Suppose L is any NP-complete language, and $R : \{0,1\}^{n \times t} \to \{0,1\}^{.01t}$ is a PPT reduction for $\mathrm{OR}_{=}(L)^t$ with success prob. $\geq .99$, target language L.

- Then, R is also a PPT reduction for $\mathrm{AND}_{=}(\overline{L})$, target language $\overline{L}$!

The modified reduction

Applying our main reduction to $\overline{L}$ in place of L, we get a reduction Q' computable by poly(n)-sized circuits:

- **Input:** $x \in \{0,1\}^n$.
- **Output:** a pair of sampling-circuit descriptions

$$\langle C \ , \ C'_x \rangle$$

where:
  - C samples from $R(\overline{D}^*)$,
  - C' samples from $R(\overline{D}^*[x;\mathbf{j}])$, $\quad \mathbf{j} \in_r [t]$.

- **New property:** if $x \in L_n$, then

$$||C - C'_x|| \ \geq \ .98 \ ,$$

while if $x \in \overline{L}_n$,

$$||C - C'_x|| \ \leq \ .3 \ .$$

The modified reduction

- Finally, we run the Proof-of-closeness proof system on the output $\langle C, C'_x \rangle$ to be convinced that the two distributions are close, i.e., that we are in Case (ii) of $\mathrm{SD}^{\geq .9}_{\leq .3}$, i.e., $x \in \overline{L}_n$.
  Gives an interactive proof for $\overline{L}$.

- Again we find $\overline{L} \in \mathrm{NP/poly}$, so again we conclude

$$\mathrm{NP} \subset \mathrm{coNP/poly} .$$

  So if L is NP-complete, the compression reduction R we assumed for $\mathrm{OR}_{=}(L)$ (with two-sided error) is unlikely to exist.

The statistical distance problem $\mathrm{SD}^{\geq .9}_{\leq .3}$

Problem (SD)

- **Input:** *sampling-circuits* $\langle C, C' \rangle$.
- **Distinguish:** *Case (i):* $||C - C'|| \geq .9$;
  Case (ii): $||C - C'|| \leq .3$.

- In instances output by our reduction Q described earlier, derived from the compression reduction R for $\mathrm{AND}_{=}(L)$, the first circuit $C = R(\overline{\mathcal{D}}^{*})$ depends only on the input length n!
- Using non-uniformity, we can give a much simpler proof system to witness Case (ii) in this special case, without using [Fortnow '87, Sahai-Vadhan '99] (this is unpublished work)

The statistical distance problem with fixed sequence

Problem (SD problem, fixed sequence)

- *Defining data:* *A non-uniform sequence* $\{C_n\}$ *of sampling circuits,* $\text{size}(C_n) \leq \text{poly}(n)$
- **Input:** *a sampling-circuit* $\langle C' \rangle$ *(of the same size as* C_n*).*
- **Distinguish:** *Case (i) :* $||C_n - C'|| \geq .9$*;*

  Case (ii): $||C_n - C'|| \leq .3$*.*

The statistical distance problem with fixed sequence

Proof system idea:

- For a given sequence $(z^1, \ldots, z^m)$ of outputs by C_n, let $\mu(z^i) := \Pr[C_n \to z^i]$. Let $\mu'(z^i) := \Pr[C' \to z^i]$.
- If $||C_n - C'|| \geq .9$ then, for most values $z^i \leftarrow C_n$,

$$\mu'(z^i) \ < .5 \cdot \mu(z^i) \ . \tag{1}$$

- If $||C_n - C'|| \leq .3$ then, for most values $z^i \leftarrow C_n$,

$$\mu'(z^i) \ > .6 \cdot \mu(z^i) \ . \tag{2}$$

- We can non-uniformly fix a $\text{poly}(n)$-sized list $z^1, \ldots, z^m$ such that:
  1. Eq. (1) holds for most z^i, for every C' in Case (ii);
  2. Eq. (2) holds for most z^i, for every C' in Case (i).

The statistical distance problem with fixed sequence

- Given C', can use **Goldwasser-Sipser set-size protocol** to prove Eq. (1) holds for most z^i. Just need $\{\mu(z^i)\}$ as non-uniform advice.

Takeaway

- We've seen new, stronger barriers to kernelization under the assumption $\mathrm{NP} \not\subseteq \mathrm{coNP/poly}$.
- Built a **non-uniform proof system** for any $\overline{L}$ for which $\mathrm{AND}_{=}(L)$ is compressible. Improved results for the case when $\mathrm{OR}_{=}(L)$ is compressible too.
- We saw that **probabilistic interaction with provers** gives a rich framework for building proof systems.
- The compression property of our AND-reduction R was used as an **information bottleneck** to fool a lying prover.
- When building our non-uniform advice, **minimax theorem** allowed us to consider probabilistic experiments, where bottleneck could be **quantified** using entropy arguments.

Thanks!