

Transformations which preserve computable randomness

and other applications of computable conditional probability

Jason Rute

University of Hawaii–Manoa

(Soon to be at Penn State)

Applications of Randomness and Analysis
June 2013

Logic and probability

Question

What do logic and probability have in common?

Answer

Information.

Question

How is such information “learned”?

Answer

Functions: measurements, surveys, computing some properties, etc.

My message

Primary Message

There are two basic types of maps:

- 1 Computable-ish maps
- 2 Computable-ish maps with a computable-ish conditional probability

The second behaves much better than the first!

Secondary Message

Computable randomness is preserved by the second type, but not always the first.

Tertiary Message

These nice maps are useful for other things, e.g. randomness reducibility.

Computable randomness

$x \in 2^{\mathbb{N}}$ is computably random if

one cannot win arbitrarily large amounts of money by betting on it

Do these maps preserve computable randomness?

For $T : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$, does it hold that

x is computably random $\Rightarrow T(x)$ is computably random?

- 1 $T(x_0x_1\dots) = x_0x_1\dots$ is the identity map. ✓
- 2 $T(x_0x_1\dots) = x_0x_2\dots$ selects the even bits. ✓
- 3 $T(x_0x_1\dots) = x_{\sigma(1)}x_{\sigma(2)}\dots$ is a computable permutation of the bits. ✓
- 4 $T(x_0x_1\dots) = x_{f(1)}x_{f(2)}\dots$ is a computable (injective) selection of the bits.
 - If the range of f is computable: ✓
 - If the range of f is not computable: **Not always! (Kasterman-Lempp)**

Question

What is common between the maps that preserve computable randomness?

Answer

The conditional probability is computable.

Notation

$$T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$$

- Ω is a Polish sample space
 - may assume it is $2^{\mathbb{N}}$.
- \mathbf{P} is a Borel probability measure on Ω
 - may assume it is fair-coin measure.
- \mathbb{X} is a Polish space
 - may assume it is $2^{\mathbb{N}}$.
- $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$ is a measurable map (random variable) from Ω to \mathbb{X} .
 - The \mathbf{P} will become significant soon.

Notation

Definition

Given $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$, define \mathbf{P}_T as the measure on \mathbb{X} given by

$$\mathbf{P}_T(B) = \mathbf{P}\{T \in B\} = \mathbf{P}(T^{-1}(B)).$$

This is called the **probability distribution of T** , or the **push-forward measure of T along \mathbf{P}** .

This is the unique measure on \mathbb{X} which makes T measure preserving.

I will often write

$$T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T).$$

Randomness preservation

Definition

A randomness notion is said to satisfy **randomness preservation** if for all total computable maps $T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$,

$$\omega \text{ is random on } (\Omega, \mathbf{P}) \quad \Rightarrow \quad T(\omega) \text{ is random on } (\mathbb{X}, \mathbf{P}_T).$$

Theorem (Folklore)

Kurtz randomness, Schnorr randomness, Martin-Löf randomness, difference randomness, weak n -randomness, n -randomness all satisfy randomness preservation.

Theorem (Rute; Bienvenu-Porter)

Computable randomness does not satisfy randomness preservation!

Theorem (Rute)

Maps with **computable conditional probability** preserve comp. randomness.

Reverse randomness preservation (no randomness ex nihilo)

Definition

A randomness notion is said to satisfy **reverse randomness preservation** if for all total computable maps $T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$,

$$x \text{ is random on } (\mathbb{X}, \mathbf{P}_T) \quad \Rightarrow \quad \exists \omega \text{ random on } (\Omega, \mathbf{P}) \text{ such that } T(\omega) = x.$$

Theorem (Shen and the Russians via Bienvenu-Porter)

Martin-Löf randomness satisfies reverse randomness preservation.

Question (Bienvenu-Porter)

Does Schnorr or comp. randomness satisfy reverse randomness preservation?

Partial Answer (Rute)

Yes ... assuming the maps have **computable conditional probability**.

Conditional Probability

Let A be an event (set).

The main idea

$\mathbf{P}[A | T]$ is the probability that A holds given knowledge of T .

Since T is a function (random variable), then so is $\mathbf{P}[A | T]$.

Actually two functions...

- $x \mapsto \mathbf{P}[A | T = x]$ is a function $\mathbb{X} \rightarrow [0,1]$.
- $\omega \mapsto \mathbf{P}[A | T](\omega)$ is a function $\Omega \rightarrow [0,1]$.

Connected by

$$\mathbf{P}[A | T](\omega) = \mathbf{P}[A | T = T(\omega)].$$

A definition of conditional probability

Assume $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$.

Discrete Case

If $\mathbb{X} = \{x_0, \dots, x_n\}$, then for $x_i \in \mathbb{X}$.

$$\mathbf{P}[A \mid T = x_i] = \frac{\mathbf{P}(A \cap \{T = x_i\})}{\mathbf{P}(\{T = x_i\})} = \frac{\mathbf{P}(A \cap T^{-1}(x_i))}{\mathbf{P}(T^{-1}(x_i))}.$$

Infinite Case

If $\mathbb{X} = 2^{\mathbb{N}}$ and $T_n : (\Omega, \mathbf{P}) \rightarrow \{0, 1\}$ is the n th bit of T , then for $x = (x_i)_{i \in \mathbb{N}} \in 2^{\mathbb{N}}$,

$$\mathbf{P}[A \mid T = x] = \lim_{n \rightarrow \infty} \mathbf{P}[A \mid (T_0, \dots, T_{n-1}) = (x_0, \dots, x_{n-1})].$$

(The case for arbitrary \mathbb{X} is similar.)

There is another definition which is more axiomatic.

Example: Conditional probability of selecting bits

Consider $E : (2^{\mathbb{N}}, \text{fair-coin}) \rightarrow (2^{\mathbb{N}}, \text{fair-coin})$ where E chooses the even bits:

$$E(\omega_0\omega_1\omega_2\omega_3\dots) = \omega_0\omega_2\omega_4\omega_6\dots$$

Then

$$\mathbf{P}[0\text{th bit of } \omega \text{ is } 1 \mid E = y] = y_0$$

$$\mathbf{P}[1\text{st bit of } \omega \text{ is } 1 \mid E = y] = \frac{1}{2}$$

$$\mathbf{P}[2\text{nd bit of } \omega \text{ is } 1 \mid E = y] = y_1$$

$$\vdots$$

$\mathbf{P}[\cdot \mid E = y]$ is the probability measure on $2^{\mathbb{N}}$ where

- the odd bits are fair-coin flips and
- the even bits are the bits of y .

Example: Projection on a product measure

Assume $(\Omega, \mathbf{P}) = (\mathbb{X} \times \mathbb{Y}, \mathbf{Q}_1 \otimes \mathbf{Q}_2)$ and $\pi_1 : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{Q}_1)$ is the projection onto the first coordinate.

Then $\mathbf{P}[A \mid \pi = x] = \mathbf{Q}_2(A_x)$ where $A_x = \{y \in \mathbb{Y} \mid (x, y) \in A\}$.

Computable notation

$$T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$$

- Ω is a **computable** Polish sample space
 - may assume it is $2^{\mathbb{N}}$.

- \mathbf{P} is a **computable** Borel probability measure on Ω
 - may assume it is fair-coin measure.

- \mathbb{X} is a **computable** Polish space
 - may assume it is $2^{\mathbb{N}}$.

- $T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$ is an **effectively** measurable map (random variable) from Ω to \mathbb{X} .
 - And \mathbf{P}_T will turn out to be computable from \mathbf{P} and T .

Effectively measurable maps

Definition

$T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$ is **effectively measurable** if it is a comp. point in the metric

$$\rho(T, S) = \int \min\{d_{\mathbb{X}}(T(\omega), S(\omega)), 1\} d\mathbf{P}(\omega) \quad (d_{\mathbb{X}} \text{ is the metric on } \mathbb{X}).$$

In the above definition, T is an equivalence class. However, we can use “Schnorr layerwise computability” to define $T(\omega)$ for all Schnorr ω .

Definition

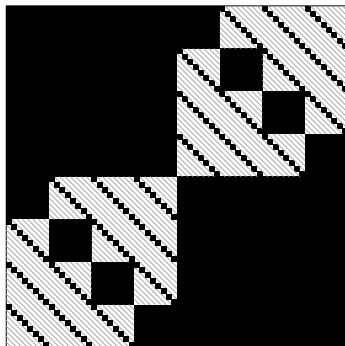
$T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$ is **Schnorr layerwise computable** if there is a Schnorr test (U_n) such that T is uniformly computable on $\Omega \setminus U_n^c$ for all n .

Theorem

- T is effectively meas. iff it is a.e. equal to a Schnorr layerwise comp. T' .
- $T(\omega)$ is well-defined for Schnorr rand. ω and Schnorr layerwise comp. T

Example

Consider this $T : ([0,1]^2, \text{Lebesgue}) \rightarrow \{\text{Black, White}\}$.



- It is effectively measurable/Schnorr layerwise computable
- But it is not computable—not even on a measure one set!
 - Black is dense in white, but white has positive probability.

L^1 -computable maps

Definition

$f : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ is **L^1 -computable** if it is a comp. point in the metric

$$\rho(f, g) = \int |f - g| d\mathbf{P}.$$

Theorem

- T is L^1 -computable iff T is effectively measurable and has comp. L^1 -norm.
- T is L^1 -computable iff T is a.e. equal to a Schnorr layerwise computable function with a computable L^1 -norm.

Computable conditional probability

We are interested in maps $T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$ such that

- T is effectively measurable
- $x \mapsto \mathbf{P}[\cdot \mid T = x]$ is effectively measurable $(\mathbb{X}, \mathbf{P}_T) \rightarrow \text{Prob}(\Omega)$.

Call such maps **computable factor maps**.

Conditional probability is not computable

Theorem (Ackerman-Freer-Roy; compare with Yu, Hoyrup-Rojas-Weihrauch)

$\mathbf{P}[\cdot | T]$ is not always computable in any nice sense, even for computable T .

Proof.

Take $T(\omega_0\omega_1\omega_2\dots) = \omega_{f(0)}\omega_{f(1)}\omega_{f(2)}\dots$ where f is injective and has range the halting problem. \square

Preservation of computable randomness

Theorem (Rute)

Computable randomness is preserved by computable factor maps.

Definition of computable randomness on $(2^{\mathbb{N}}, \mathbf{P})$

Definition

On $(2^{\mathbb{N}}, \mathbf{P})$, a **computable dyadic martingale** is a partial computable function $M: \{0,1\}^* \rightarrow \mathbb{R}^+$ such that

- 1 $M(\sigma 0)\mathbf{P}(\sigma 0) + M(\sigma 1)\mathbf{P}(\sigma 1) = M(\sigma)\mathbf{P}(\sigma)$
- 2 The measure $\mathbf{Q}(\sigma) = M(\sigma)\mathbf{P}(\sigma)$ is total computable.
(Here *undefined* $\cdot 0 = 0$, so $M(\sigma)$ can be undefined only when $\mathbf{P}(\sigma) = 0$.)

Definition

On $(2^{\mathbb{N}}, \mathbf{P})$, x is **computably random** if neither of these two hold:

- 1 $x \in [\sigma]$ where $\mathbf{P}(\sigma) = 0$,
- 2 $\limsup_n M(\sigma) = \infty$ for some computable dyadic martingale.

This definition can be extended to other metric spaces.

Four ways to view $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$

$$1 \quad \omega \mapsto T(\omega) \quad : \quad (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$$

$$2 \quad B \mapsto T^{-1}(B) \quad : \quad \text{MeasSets}(\mathbb{X}) \rightarrow \text{MeasSets}(\Omega)$$

$$3 \quad f \mapsto f \circ T \quad : \quad \text{MeasFuns}(\mathbb{X} \rightarrow \mathbb{R}) \rightarrow \text{MeasFuns}(\Omega \rightarrow \mathbb{R})$$

$$4 \quad \mathbf{Q} \mapsto \mathbf{Q}_T \quad : \quad \text{Prob}(\mathbb{X}) \ll \mathbf{P} \rightarrow \text{Prob}(\Omega) \ll \mathbf{P}_T$$

All of these maps are “computable” for effectively measurable T .

Four reverse ways to view $T : (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$

$$1 \quad x \mapsto \mathbf{P}[\cdot \mid T = x] \quad : \quad (\mathbb{X}, \mathbf{P}_T) \rightarrow \text{Prob}(\Omega)$$

$$2 \quad A \mapsto \mathbf{P}[A \mid T = \cdot] \quad : \quad \text{MeasSets}(\Omega, \mathbf{P}) \rightarrow L^1(\Omega, \mathbf{P})$$

$$3 \quad f \mapsto \mathbf{E}[f \mid T = \cdot] \quad : \quad L^1(\Omega, \mathbf{P}) \rightarrow L^1(\mathbb{X}, \mathbf{P}_T)$$

$$4 \quad \mathbf{Q} \mapsto \mathbf{Q}^T \quad : \quad \text{Prob}(\mathbb{X}) \ll \mathbf{P}_T \rightarrow \text{Prob}(\Omega) \ll \mathbf{P}$$

If $\mathbf{Q} \ll \mathbf{P}_T$, that is $\mathbf{Q}(B) = \int_B f d\mathbf{P}_T$ for some $f \in L^1$, then

$$\mathbf{Q}^T(A) = \int_A f \circ T d\mathbf{P} = \int \mathbf{P}[A \mid T = \cdot] d\mathbf{Q}.$$

All of these maps are “computable” for computable factor maps T .

Theorem (Rute)

The following are equivalent for an effectively measurable map $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$.

- 1 $x \mapsto \mathbf{P}[\cdot \mid T = x]$ is effectively measurable from $(\mathbb{X}, \mathbf{P}_T) \rightarrow \text{Prob}(\Omega)$.
- 2 For any measure $\mathbf{Q} \ll \mathbf{P}_T$, the measure \mathbf{Q}^T is computable from \mathbf{Q} and the modulus of absolute continuity $\delta(\varepsilon)$:

$$\mathbf{Q}(B) \leq \delta(\varepsilon) \quad \Rightarrow \quad \mathbf{P}(B) \leq \varepsilon.$$

Proof sketch for randomness preservation

Proof sketch of randomness preservation for computable randomness.

- Take $T(\omega) = x$ which is not computably random on $(\mathbb{X}, \mathbf{P}_T)$.
- Take a martingale M which succeeds on x .
- Make the martingale “nice”:
 - Has the savings property: Puts all but €1 into a savings account.
 - Has the bounded increments: Never gains more than €1 at a time.
- The corresponding measure $\mathbf{Q}(\sigma) = M(\sigma)\mathbf{P}_T(\sigma)$ is absolutely continuous with modulus of continuity $\delta(\varepsilon) = \varepsilon^2/4$.
- We can compute \mathbf{Q}^T since T is a **computable factor map** and the modulus of abs. cont. for \mathbf{Q} is computable.
- The martingale N corresponding to \mathbf{Q}^T succeeds on ω .
- Hence ω is not computably random on (Ω, \mathbf{P}) .



Reverse preservation of Schnorr/computable randomness

Theorem (Rute)

Assume

- $T: (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$ is a **computable factor map**.
- $x \in (\mathbb{X}, \mathbf{P}_T)$ is Schnorr random (resp. computably random).

There is a Schnorr random (resp. comp. random) $\omega \in (\Omega, \mathbf{P})$ such that $T(\omega) = x$.

For Schnorr rand., this follows from a variation of van Lambalgen's theorem.¹

Theorem (Rute)

For a computable factor map $T: (\Omega, \mathbf{P}) \rightarrow (\mathbb{X}, \mathbf{P}_T)$,

$$\left(\begin{array}{l} \omega \text{ is Schnorr on } \mathbf{P} \\ \& x = T(\omega) \end{array} \right) \Leftrightarrow \left(\begin{array}{l} x \text{ is Schnorr on } \mathbf{P}_T \\ \& \omega \text{ is "Schnorr on } \mathbf{P}[\cdot | T = x] \text{ relative to } x" \end{array} \right).$$

¹Use uniformly relativized Schnorr randomness for non-computable measures / oracles.

A characterization of computable randomness

Let $\Omega = 2^{\mathbb{N}}$ and \mathbf{P} be the fair-coin measure. Fix \mathbb{X} .

Theorem (Rute)

*If \mathbf{Q} is a computable probability measure on \mathbb{X} ,
then $\mathbf{Q} = \mathbf{P}_T$ for some computable factor map $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$.*

Theorem (Rute)

Assume \mathbf{Q} is a computable probability measure. The following are equivalent:

- 1** $x \in \mathbb{X}$ is computably random on \mathbf{Q} .
- 2** $x = T(\omega)$ for some computably random $\omega \in \Omega$
and some computable factor map $T : (\Omega, \mathbf{P}) \rightarrow \mathbb{X}$ with $\mathbf{Q} = \mathbf{P}_T$.

Proof.

\Leftarrow : Use preservation of randomness.

\Rightarrow : Use reverse preservation of randomness. □

A new randomness reducibility

- There are many notions of randomness reducibility: e.g. $\leq_K, \geq_{LR}, \geq_{Sch}$
- The goal is to say one sequence is more random than another.
- Here is a new reducibility based on randomness preservation:

Definition

Assume x and y are Schnorr random on $2^{\mathbb{N}}$ with the fair coin measure. Define

$$x \leq_{fm} y \iff \exists T: (2^{\mathbb{N}}, \text{fair-coin}) \rightarrow (2^{\mathbb{N}}, \text{fair-coin}) \quad y = T(x)$$

where T is a measure-preserving, computable factor map.

Basic properties

- $x \leq_{fm} y$ means “ y is **more random** than x ”.
- By randomness preservation, Schnorr, computable, and Martin-Löf randomness are all upward closed in \leq_{fm}
- $\leq_{fm} \Rightarrow \geq_{tt}, \geq_{Sch}, \geq_{LR}$.

Randomness reducibility on other measures

This can be extended to other measures as well.

Definition

Assume

- \mathbf{P}, \mathbf{Q} are computable probability measures on $2^{\mathbb{N}}$.
- x and y are Schnorr random, resp., on $(2^{\mathbb{N}}, \mathbf{Q})$ and $(2^{\mathbb{N}}, \mathbf{P})$.

Define

$$(x, \mathbf{P}) \leq_{fm} (y, \mathbf{Q}) \iff \exists T: (2^{\mathbb{N}}, \mathbf{P}) \rightarrow (2^{\mathbb{N}}, \mathbf{Q}) \quad y = T(x)$$

where T is a measure preserving, computable factor map.

- (x, \mathbf{P}) is \leq_{fm} -maximum (most random) $\iff x$ is computable and $\mathbf{P} = \delta_x$.
- $(x, \mathbf{P}) \leq_{fm} (y, \mathbf{Q}) \implies x \geq_{tt} y$, and " $\mathbf{P} \geq_{tt} \mathbf{Q}$ ".
- \leq_{fm} can even be extended to noncomputable measures.

Computable sub- σ -algebras

Let \mathcal{B} be the Borel sigma algebra on $(\Omega, \mathcal{B}, \mathbf{P})$.

Definition

$\mathcal{F} \subseteq \mathcal{B}$ is a **computable sub- σ -algebra** if this map is effectively measurable

$$\omega \mapsto \mathbf{P}[\cdot \mid \mathcal{F}](\omega).$$

Theorem

The following are equivalent.

- 1 \mathcal{F} is a computable sub- σ -algebra.
- 2 $\mathcal{F} = \sigma(T)$ for some computable factor map T .

Remark

Say that $\mathcal{F} \subseteq \mathcal{B}$ is a c.e. sub- σ -algebra if $\mathcal{F} = \sigma(T)$ for some eff meas map T . This also has nice properties, but the one above is better.

Convergence: ergodic theorem

Definition

The space (Ω, \mathbf{P}, T) is a **measure preserving system (f.p.s.)** if T is measure preserving.

It is **computable** if T is effectively measurable.

Define $\mathcal{I}nv(T) = \{A : T^{-1}(A) = A\}$ as sub- σ -algebra of T -invariant sets.

Theorem (Hoyrup, Rojas, Gács)

Let (Ω, \mathbf{P}, T) be a comp. m.p.s. and let f be L^1 -comp. If

- 1** \mathcal{I} is computable (\Leftrightarrow *the ergodic decomposition is computable*)
(Note, (Ω, \mathbf{P}, T) is ergodic $\Leftrightarrow \mathbf{P}[\cdot \mid \mathcal{I}nv(T)] = \mathbf{P}$)

then $\frac{1}{n} \sum_{k < n} (f \circ T^k)(\omega) \rightarrow \mathbf{E}[f \mid \mathcal{I}nv(T)](\omega)$ on Schnorr randoms ω .

This is tight. Without **1**, need Martin-Löf randomness (Franklin-Towsner).

Convergence: Levy 0-1 law

Definition

The space $(\Omega, \mathbf{P}, (\mathcal{F}_n)_{n \in \mathbb{Z}})$ is a **filtered probability space (f.p.s.)** if

$$\cdots \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$$

It is **computable** if $(\mathcal{F}_n)_{n \in \mathbb{Z}}$ is a computable sequence of sub- σ -algebras. Define $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ and $\mathcal{F}_{\infty} = \sigma(\bigcup_n \mathcal{F}_n)$.

Theorem (Rute)

Let $(\Omega, \mathbf{P}, (\mathcal{F}_n)_{n \in \mathbb{Z}})$ be a comp. f.p.s. and let f be L^1 -computable. If

1 $\mathcal{F}_{\pm\infty}$ is computable

then $\mathbf{E}[f \mid \mathcal{F}_n](\omega) \xrightarrow[n \rightarrow \pm\infty]{} \mathbf{E}[f \mid \mathcal{F}_{\pm\infty}](\omega)$ on Schnorr randoms ω .

This is tight. Without **1**, probably need Martin-Löf randomness.

Convergence: Doob martingale convergence theorem

Definition

A **martingale** on $(\Omega, \mathbf{P}, (\mathcal{F}_n)_{n \in \mathbb{N}})$ is a sequence of integrable functions:

- M_n is \mathcal{F}_n -measurable.
- $\mathbf{E}[M_n \mid \mathcal{F}_m] = M_m$ for $n \geq m$.

Theorem (Rute)

Let (M_n) be a martingale on (Ω, \mathbf{P}) such that

- 1 M_n is L^1 -computable,
- 2 $\sup \|M_n\|_{L^1}$ is a computable real,
- 3 $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$ is computable.

Then $M_n(x)$ converges on computable randoms.

This is probably tight. Without 3, need Martin-Löf randomness.

Questions

Does reverse randomness preservation (no randomness ex nihilo) hold for Schnorr and computable randomness for all effectively measurable maps?

Final remark

I hope I convinced you that there is a **deep connection** between

- information in probability theory (conditioning)
- information in randomness (computability)

Thank You!

These slides will be available on my webpage:

math.cmu.edu/~jrute

Or just Google™ me, “Jason Rute”.