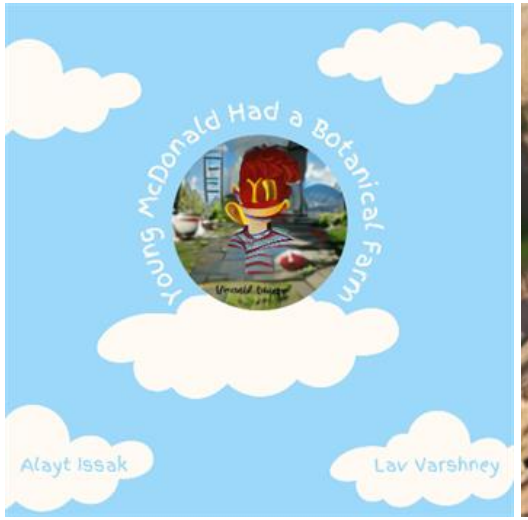
Generative AI Models
ECE 598 LV – Lecture 23

Lav R. Varshney

14 April 2022

- Mathematical theories of creativity
- Generative and creative AI models for art

AI-generated art, music, culinary, etc.




PRINTS AND MULTIPLES

"A penguin on Mars wearing a spacesuit walking a robot dog next to Santa Claus”

CLIP+VQGAN

https://twitter.com/mark_riedl/status/1511779188486852613

"A penguin on Mars wearing a spacesuit walking a robot dog next to Santa Claus”

DALL-E-2

"A penguin on Mars wearing a spacesuit walking a robot dog next to Santa Claus”

DALL-E-2 (with some more prompting techniques)

"A penguin”

CLIP+VQGAN

"A penguin on Mars”

CLIP+VQGAN

"A penguin on Mars wearing a spacesuit”

CLIP+VQGAN

"A penguin on Mars wearing a spacesuit and walking a dog”

CLIP+VQGAN

AI Music

https://magenta.tensorflow.org/music-transformer

https://www.youtube.com/watch?v=Emidxpkyk6o

Mona Lisa Smile

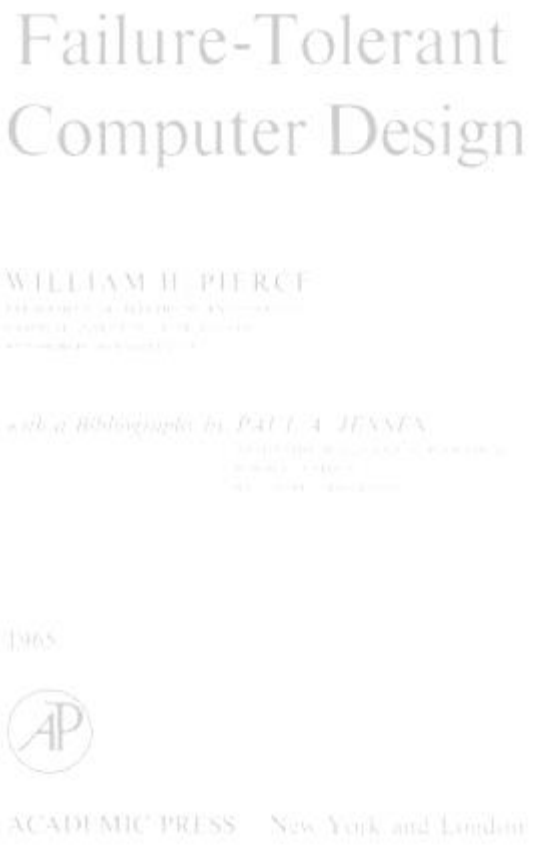
https://www.youtube.com/watch?v=7bLUG4g0cVc

https://www.youtube.com/watch?v=-2RqJIXFpYw

Intentionality in Art (Communicative Intent)

- Especially in the Western tradition following Romanticism, communication of meaning in art is necessary for eliciting an aesthetic experience (Csikszentmihalyi and Robinson 1990; Cilliers 1998; Ritchie 2007)
  - See criticisms, e.g. (Wimsatt and Beardsley 1946)
- Considering narration or poetry, (linguistic) meaning is the relation between a linguistic form and communicative intent, where communicative intents are about things that are outside of language. Communicative intent is distinct from standing meaning, which is constant across all of its possible contexts of use (Bender and Koller 2020).
- Recent surveys further indicate people want not just novelty/quality, but also intentionality and autonomy, to attribute creativity to an artificial system (Ventura 2019)

“what every engineer needs is a good set of limit theorems”



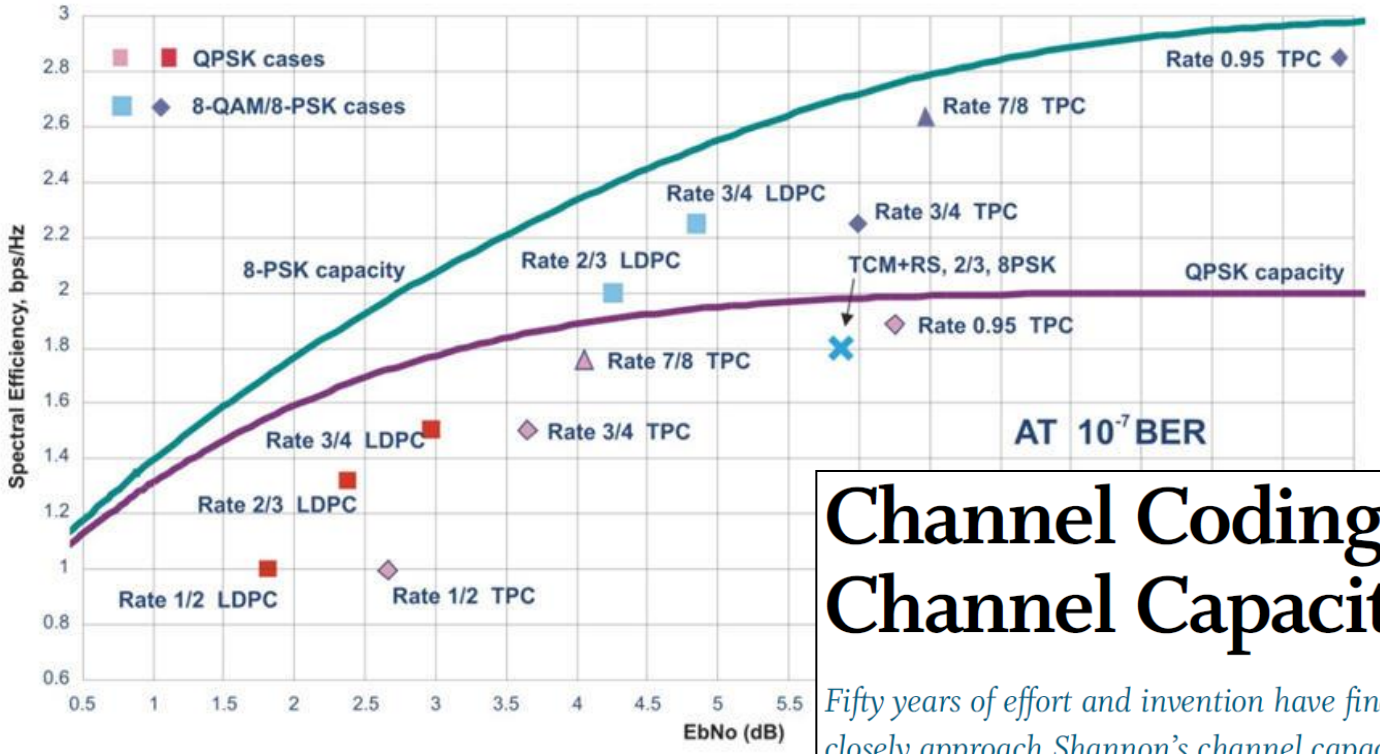
"what every engineer needs is a good set of limit theorems"

Why?

Fundamental Limits

- Fundamental benchmarks that allow an evaluation of new technologies on an absolute scale, rather than only compared to previous technologies
- Ideals for pushing industry and researchers to build technologies that approach/achieve these limits
- Establish the playing field in terms of which resources and performance criteria are fundamental and which are largely unimportant
- In delineating what is possible from what is impossible, provide insights into operating at the boundary, i.e. principles for optimal designs

The Quest for Shannon Capacity



Channel Coding: The Road to Channel Capacity

Fifty years of effort and invention have finally produced coding schemes that closely approach Shannon's channel capacity limit on memoryless communication channels.

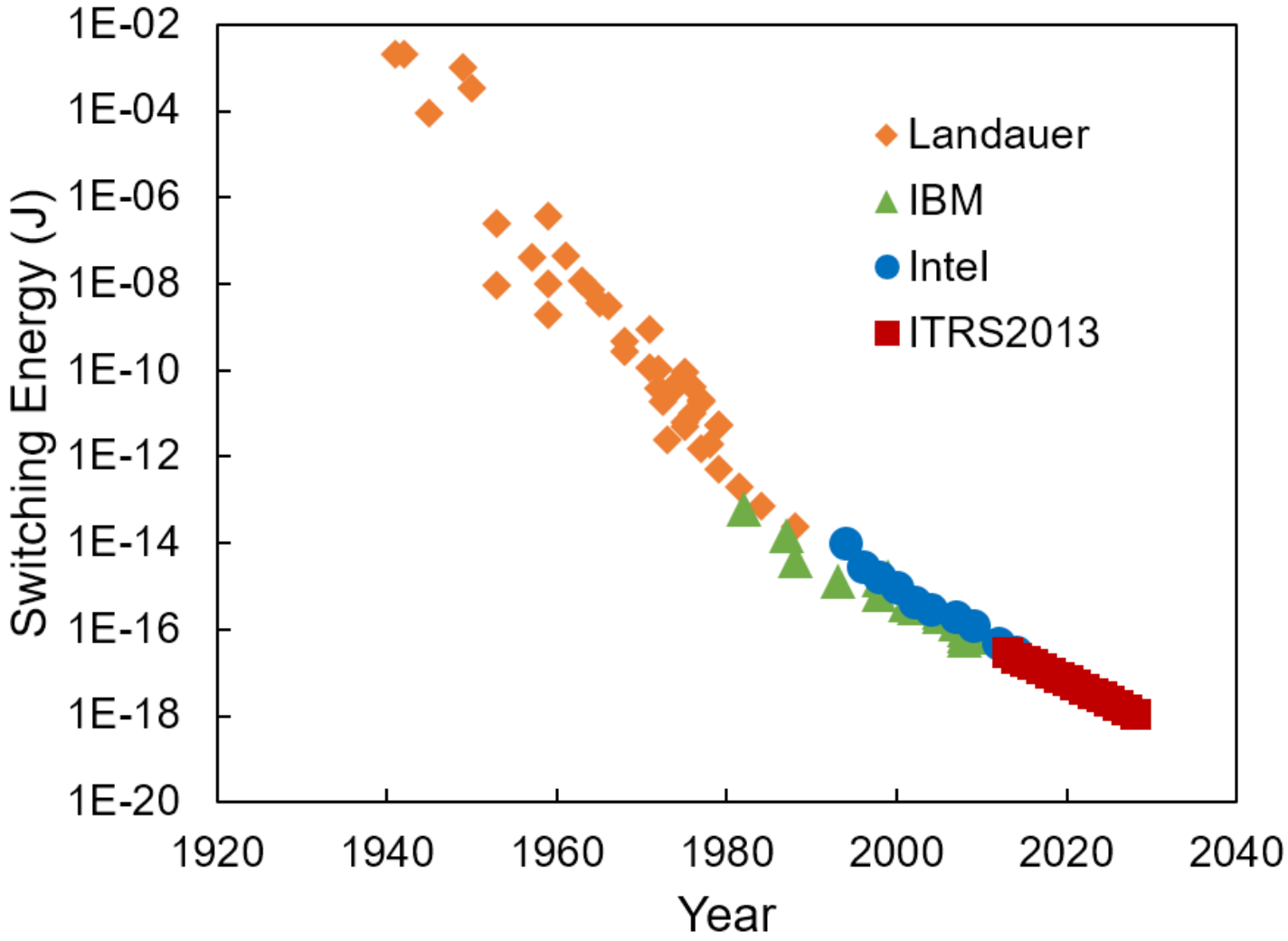
By Daniel J. Costello, Jr., *Fellow IEEE*, and G. David Forney, Jr., *Life Fellow IEEE*

IEEE COMMUNICATIONS LETTERS, VOL. 5, NO. 2, FEBRUARY 2001

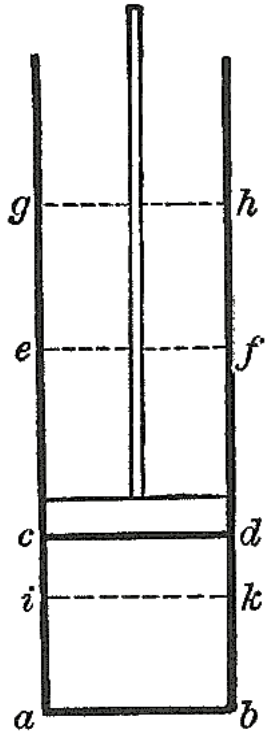
On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit

Sae-Young Chung, *Member, IEEE*, G. David Forney, Jr., *Fellow, IEEE*, Thomas J. Richardson, and Rüdiger Urbanke

The Quest for Semiconductor Scaling



[H.-S. P. Wong, C.-S. Lee, and J. Luo, "CMOS Technology Scaling Trend," https://nano.stanford.edu/cmos-technology-scaling-trend, accessed Aug. 2017.]



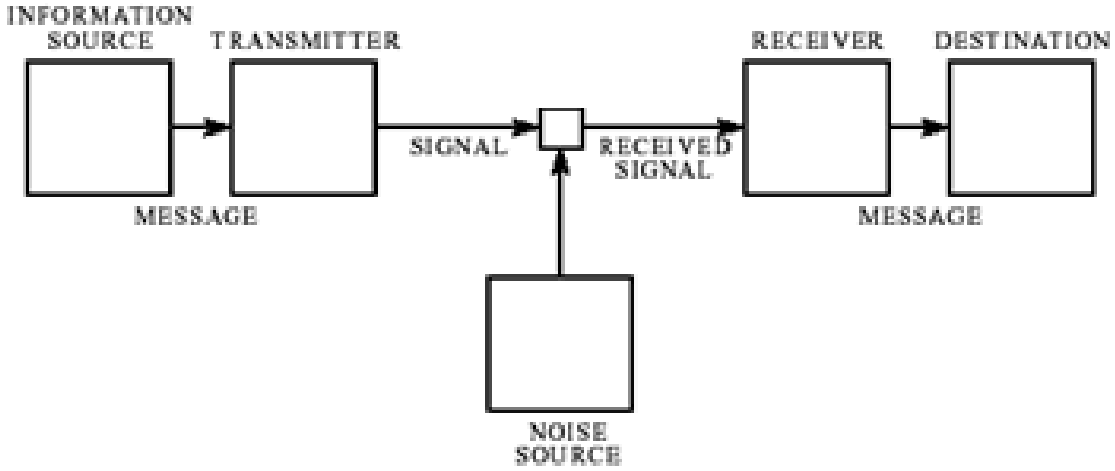
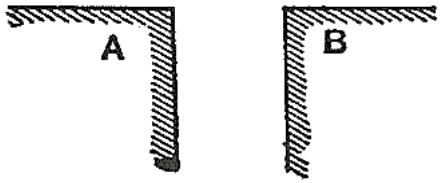
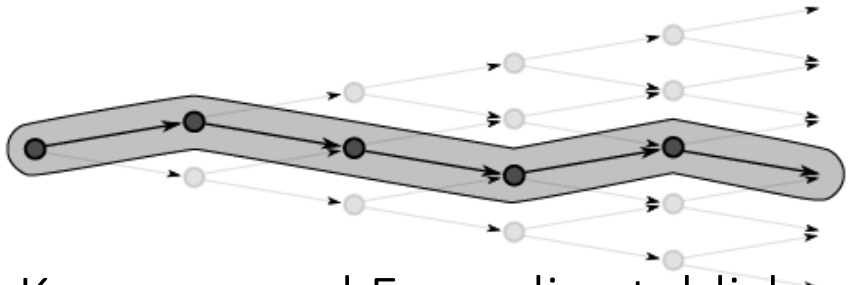


Fig. 1 — Schematic diagram of a general communication system.

Shannon established fundamental limits of communication in the presence of noise



Carnot established fundamental limits on efficiency of engines



Karaman and Frazzoli established a fundamental speed limit of flight in forests without crashing

INVITED PAPER

Shannon-Inspired Statistical Computing for the Nanoscale Era

By NARESH R. SHANBHAG, *Fellow IEEE*, NAVEEN VERMA, *Member IEEE*, YONGJUNE KIM, *Member IEEE*, AMEYA D. PATIL, *Student Member IEEE*, AND LAV R. VARSHNEY, *Senior Member IEEE*

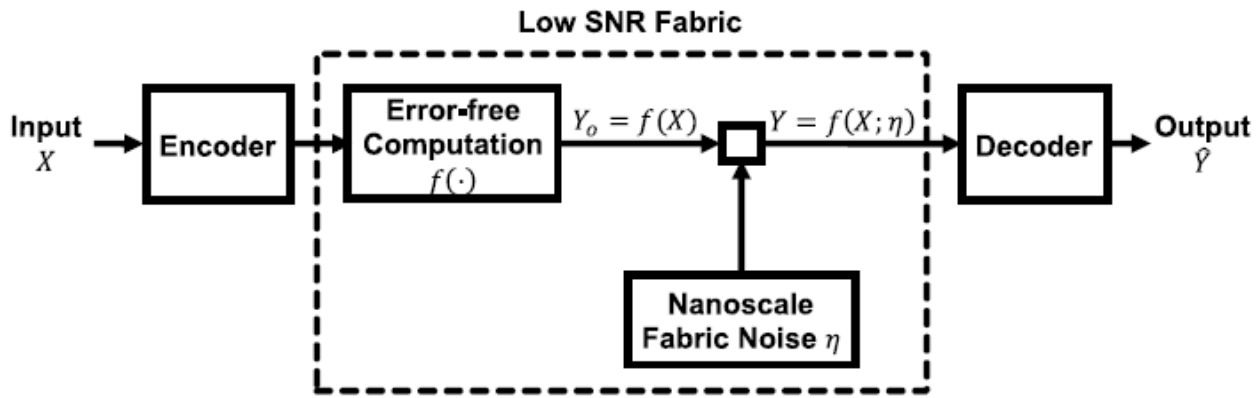


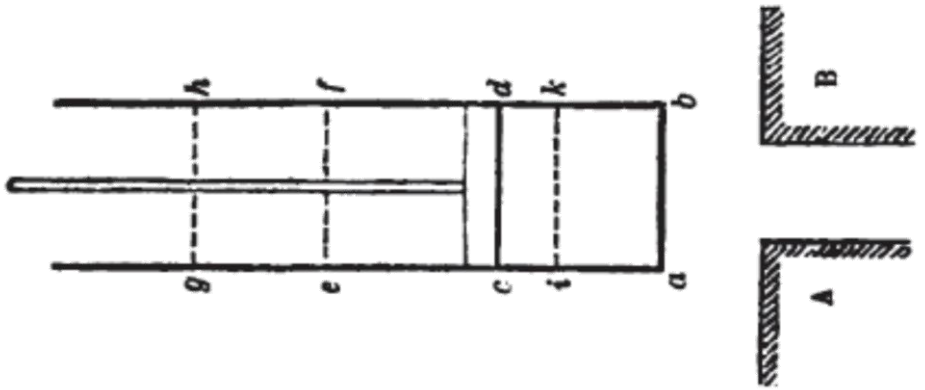
Fig. 2. ***Shannon-inspired statistical model of computation.***

Defining a Closed Deductive System

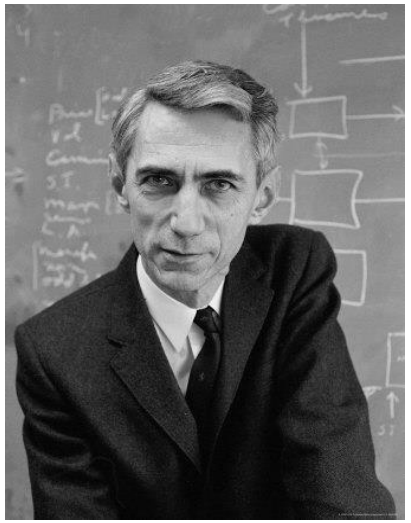
Carnot limit

"Carnot's know-how is not about building a machine but rather a diagram. This diagram is drawn in such a way that it allows one to move from one engine to any other, and indeed to nonexistent engines simply drawn on paper. Real three-dimensional steam engines are interesting but localized and cumbersome."

- B. Latour (1990)



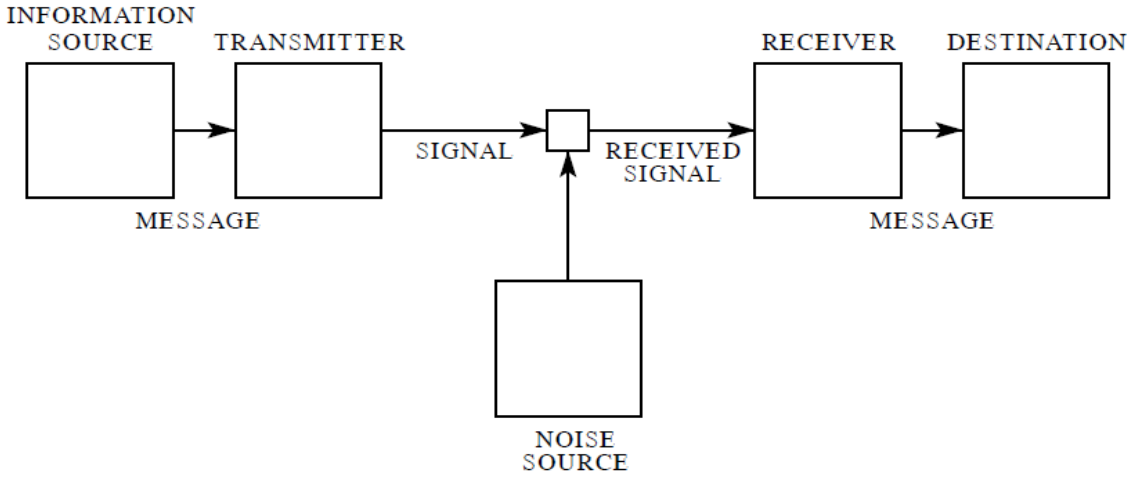
Defining a Closed Deductive System



Shannon capacity

"Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem."

- C. E. Shannon (1948)

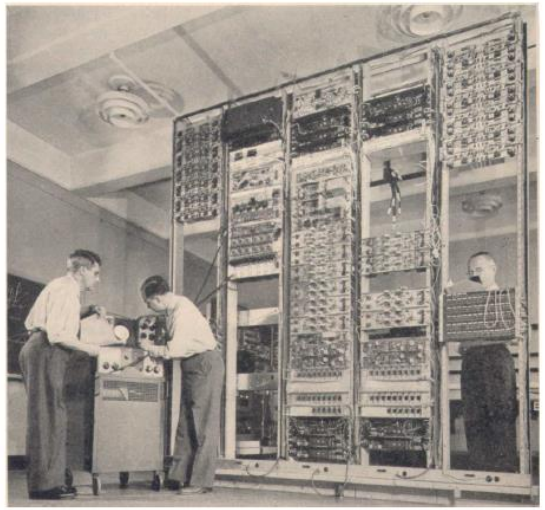


Shannon a la Gallager

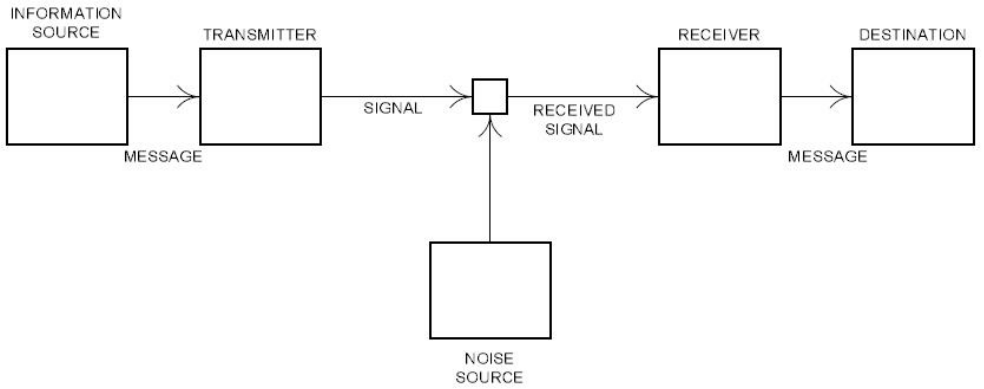
- Shannon was almost opposite of applied mathematicians
  - Applied mathematicians solve mathematical models formulated by others (perhaps with minor changes to suit their tools)
- Shannon was a creator of models — his genius lay in determining the core of the problem and removing details that could be reinserted later

- Shannon asked conceptual questions about everyday things

Fundamental Resources and Metrics



MODEL



In moving from what Galileo called *machines in the concrete* to *machines in the abstract*, one necessarily restricts attention to the fundamental resources (e.g. time, energy, volume, knowledge) and fundamental performance metrics (e.g. reliable computation rate), while ignoring certain other resources and metrics.

Memory
Deductive reasoning
Association
Perception
Introspection
Abductive reasoning
Inductive reasoning
Problem solving
Language
Attention
Creativity

Computational creativity systems and their performance

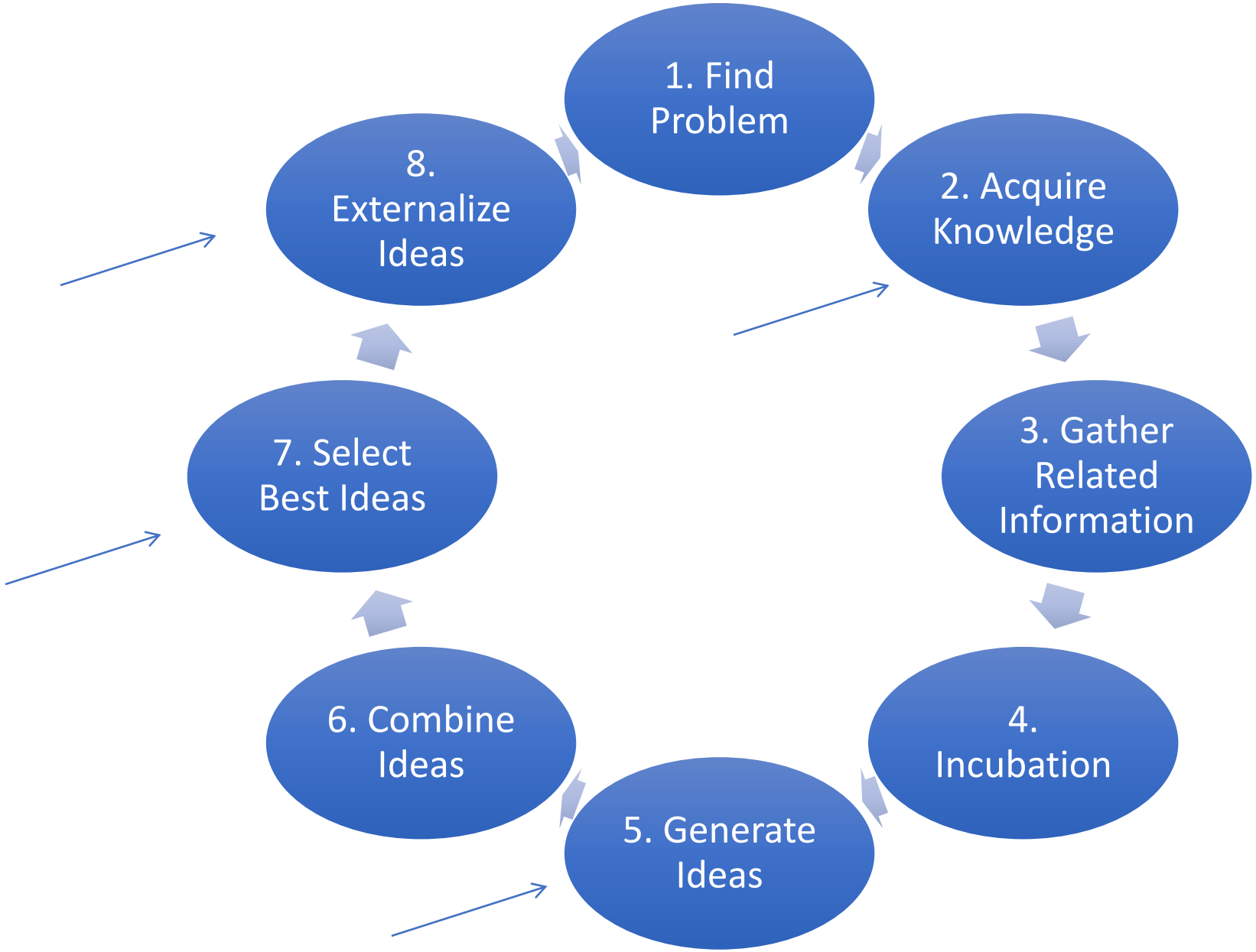
- *Computational creativity* systems typically involve combinatorial creativity, combining familiar ideas into ones that are new to the whole of human history (so-called H-creativity)
- The Turing and Lovelace 2.0 tests for intelligence make binary comparisons to either human ability or a fixed threshold of validity for any creative domain
- There is desire to score cognitive abilities in a graded manner, rather than just all-or-nothing
- Aim to define an absolute quantitative scale of performance, rather than just comparison to people or previous systems: a general theory of creativity that yields fundamental limits

Are there fundamental limits to how creative any system can be in a given domain?

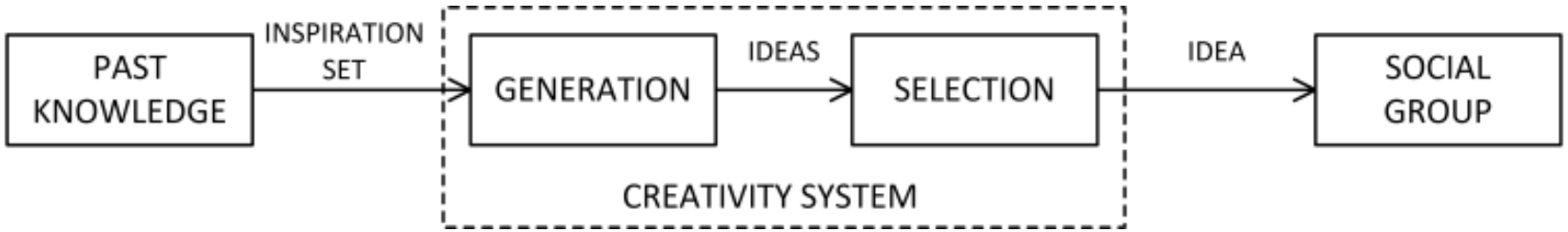
Creativity is the generation of an artifact that is judged to be novel and also to be appropriate, useful, or valuable by a suitably knowledgeable social group.

Process and intentionality are irrelevant

- Creative systems follow some process of creativity, which is only relevant for algorithmic or neuroscientific studies of creativity
- Frequently creative agents are said to have *intentionality*, a human intent, inspiration, or desire to express something
    - These aspects of intent in creativity are irrelevant to the engineering problem and we take computational creativity systems at interface value
- The significant aspect is that a creative procedure selects a creative product from a (partial) enumeration of the conceptual space of possible products in such a way that both novelty and quality (appropriateness, utility, or value) are simultaneously large

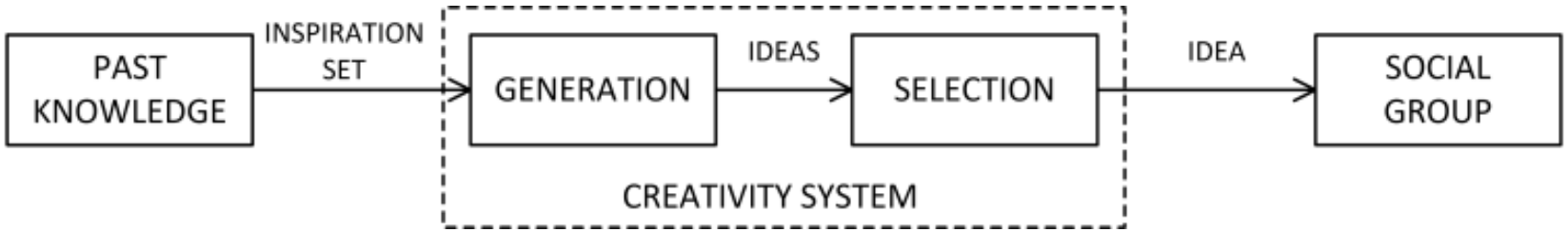

1. Find Problem
2. Acquire Knowledge
3. Gather Related Information
4. Incubation
5. Generate Ideas
6. Combine Ideas
7. Select Best Ideas
8. Externalize Ideas

[Sawyer, 2012]



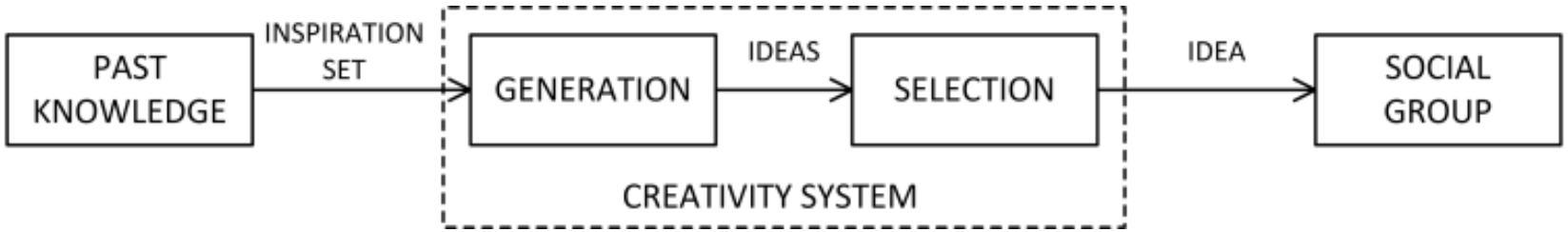
Block diagram of a general creativity system.

The *past knowledge* from the historical record of a given creative domain, which may or may not be well-organized and well-structured, is used to derive an inspiration set of ideas or artifacts that have previously been created. This past knowledge could be corpora of: recipes in culinary, compositions or recordings in music, process flow diagrams in chemical engineering, or card tricks in magic.



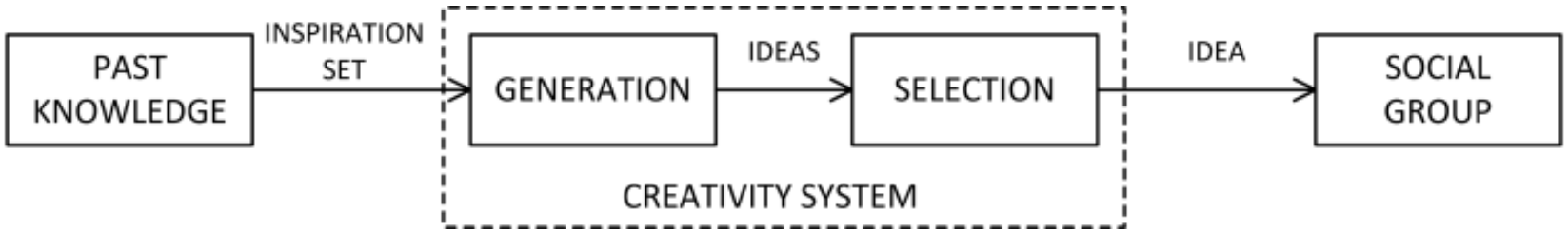
Block diagram of a general creativity system.

The subsystem for *generation* enumerates a subset of ideas or artifacts in the conceptual space for the creative domain. This could be as simple as listing the entire conceptual space in a manner suitable for selection or it could implement a complicated algorithm that intelligently generates only a handful of possibilities in the first place. The generation of ideas is often considered the key step in creativity and is rather different from other forms of reasoning such as induction or deduction.



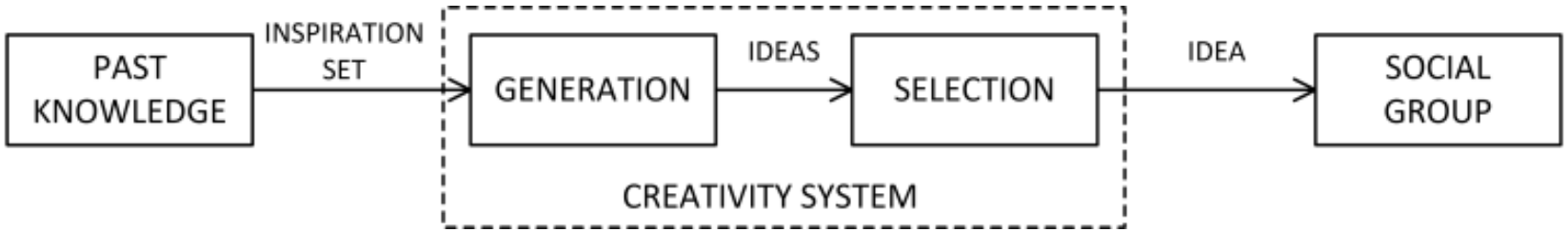
Block diagram of a general creativity system.

The subsystem for *selection* chooses one idea or artifact from the list provided by the generation subsystem, typically drawing on domain knowledge models to assess novelty and quality. This is sometimes referred to as convergent thinking, to follow the divergent thinking phase of creativity.



Block diagram of a general creativity system.

The *social group* is the audience for which the creative product is intended (and that will judge it). Different social groups may or may not respond to ideas or artifacts differently.



Block diagram of a general creativity system.

Note that the two internal parts—generation and selection—are drawn separately

One might wonder if such separation may result in any loss in performance

Towards a formalism

Creativity is the generation of an **artifact** that is judged to be **novel** and also to be appropriate, **useful**, or valuable by a suitably **knowledge**able social group.

Towards a formalism

Component An atomic unit in the creative domain, drawn from the set Ω, from which artifacts are constructed.

Discrete Artifact An unordered combinatorial object α selected from the power set 2^{Ω} of possible *components,* Ω, that define the creative domain. (assume all possible components known)

Continuous Artifact A vector α selected from the set $[0,1]^{|\Omega|} \subset \mathbb{R}_{+}^{|\Omega|}$, a space governed by the possible components, Ω, that define the creative domain.

(In certain settings, we may quantize the space of continuous artifacts into a discrete set, since small differences in artifacts may not be relevant to observers.)

Towards a formalism

Known Set Set of artifacts that are already known in the creative domain, Θ, also called the *inspiration set*.

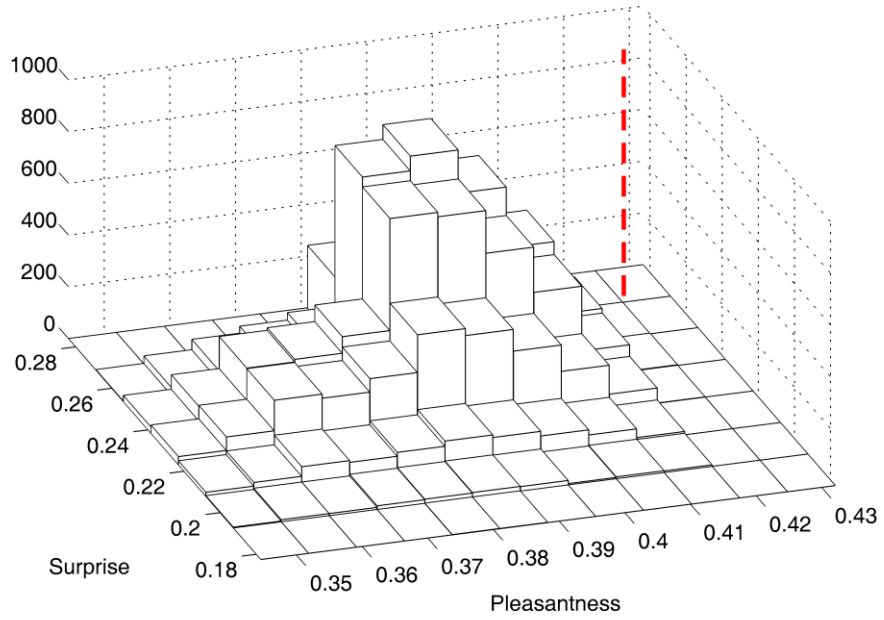
In the discrete case, $\Theta \subseteq 2^{\Omega} \in 2^{2^{\Omega}}$, whereas in the continuous case, Θ is a set of vectors in $[0,1]^{|\Omega|}$, drawn from the space of possible vector-sets $\mathcal{V}$.

Novelty A non-negative function, in the discrete case $s: 2^{\Omega} \times 2^{2^{\Omega}} \rightarrow \mathbb{R}_+$, that measures the surprise of a given artifact α_0 in the presence of a known set Θ, e.g. the empirical Bayesian surprise. In the continuous case, $s: [0,1]^{|\Omega|} \times \mathcal{V} \mapsto \mathbb{R}_+$.

Utility A non-negative function $q: 2^{\Omega} \rightarrow \mathbb{R}_+$ that measures the quality of a given discrete artifact α_0, e.g. through the psychophysical properties of components and their combining rules. Analogously $q: [0,1]^{|\Omega|} \mapsto \mathbb{R}_+$ for continuous artifacts.

Towards a formalism

- A coding scheme for channel coding may be thought of as a test source with an input distribution, for informational characterization
- Similarly, think of a creativity algorithm as a (possibly degenerate) probabilistic process $P_A(\alpha)$ for mathematical characterization
  - (stochastic sampling)



Joint histogram of surprise and pleasantness for 10000 generated Caymanian Plantain Dessert recipes. Values for the selected/tested recipe indicated with red dashed line.

Towards a formalism

Definition A *generative algorithm* $\mathcal{G}$ is a probabilistic process $P_A(\alpha)$ that produces a set of artifacts $\{\alpha_i\}_{i=1}^{n}$.

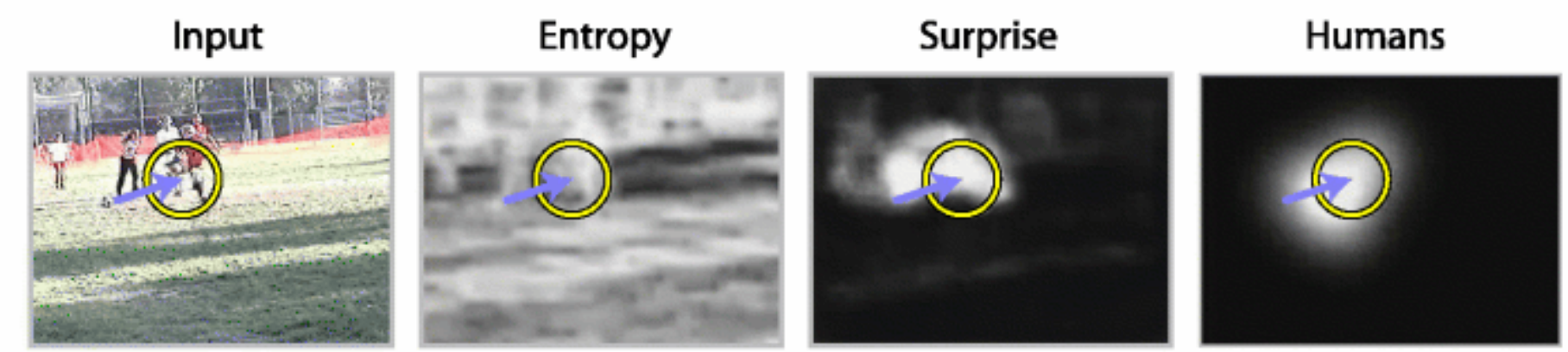
Definition A *selective algorithm* $\mathcal{S}$ chooses one member of the set $\{\alpha_i\}_{i=1}^{n}$.

Definition A *creativity algorithm* $\mathcal{C} = (\mathcal{G}, \mathcal{S})$ is the concatenation of a generative and selective algorithm.

Bayesian surprise as novelty

$$S(R,\mathcal{B}) = D\left(P_{B|R}||P_B\right) = \int_{\mathcal{B}} P_{B|R} \log \frac{P_{B|R}}{P_B} dB$$

- newly created recipe
- personalized repository of prior food experience
- posterior beliefs
- prior beliefs



[Itti and Baldi, 2006]

More on quality functions: two approaches

Definition A *trainable quality measure* for discrete artifacts is a non-negative function $q_n: \Omega^n \mapsto \mathbb{R}_+$ that measures the utility of a given artifact α_0 of n components.

Definition A *Sobolev quality measure* for continuous artifacts is a non-negative function $q_K: [0,1]^{|\Omega|} \mapsto \mathbb{R}_+$ that measures the utility of a given artifact α_0, where $q_K \in \Gamma^k$.

Here $\Gamma^k = \left\{q \in W_p^\lambda(\mathcal{X}): \left\|q^{(k)}\right\|_\infty \leq K \text{ for all } k \leq \lambda\right\}$ where the Sobolev space $W_p^\lambda(\mathcal{X})$ is the set of functions in $\mathcal{L}_p$ such that derivatives of order equal or less than λ exist and are in $\mathcal{L}_p$.

Limit Theorems

Recall our creativity system $\mathcal{C}$ has a generative algorithm $\mathcal{G}$ based on stochastic sampling $P_A(\alpha)$ and a selective algorithm $\mathcal{S}$ based on ranking and selecting. Let us first consider the optimal selective algorithm $\mathcal{S}$ for a given generative algorithm $\mathcal{G} = P_A(\alpha)$. Suppose we have multivariate ordering principle $\mathfrak{P}$ to aggregate novelty/quality.

Theorem An optimal selective algorithm $\mathcal{S}$ for artifacts $\{\alpha_i\}_{i=1}^n$ produced by any generative algorithm $\mathcal{G} = P_A(\alpha)$ finds the best artifact $\alpha \in \{\alpha_i\}_{i=1}^n$ according to ordering principle $\mathfrak{P}$.

We can focus on optimizing the stochastic sampling algorithm (and invoke the theory of concomitants of order statistics to analyze selection)

Basic Tradeoff in Creativity: Probabilities

Consider algorithm that produces novel and high-quality artifacts with probabilities above thresholds λ_s and λ_q

$$S(Q) = \max_{P_A(\alpha):\Pr[q(A)>\lambda_q]\geq Q} \Pr[s(A,\Theta) > \lambda_s]$$

Make use of information geometry techniques

Lemma [Varshney, 2013] Shannon capacity C for channel $p_{Y|X}$ is:

$$C = \min_{p_Y(y)} \max_{X} s(x)$$

Geometrically, the unconstrained optimal output distribution will be the center of a "sphere" with radius measured by Bayesian surprise, as derived from the KKT conditions

Basic Tradeoff in Creativity: Average Case

Novelty-Quality tradeoff in Creativity

$$S(Q) = \max_{P_A(\alpha):E[q(A)]\geq Q} E[s(A,\Theta)]$$

Basic Tradeoff in Creativity: Average Case

Novelty-Quality tradeoff in Creativity

$$S(Q) = \max_{P_A(\alpha):E[q(A)]\geq Q} E[s(A,\Theta)]$$

Lemma [Varshney, 2013]

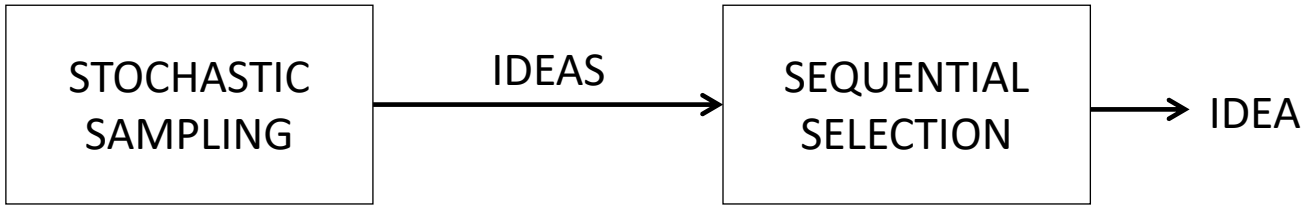
$$E[s(A,\Theta)] = I(A,\Theta).$$

Corollary

$$S(Q) = \max_{P_A(\alpha):E[q(A)]\geq Q} I(A,\Theta)$$

(Shannon's capacity-cost function)

Optimal Creativity Algorithms

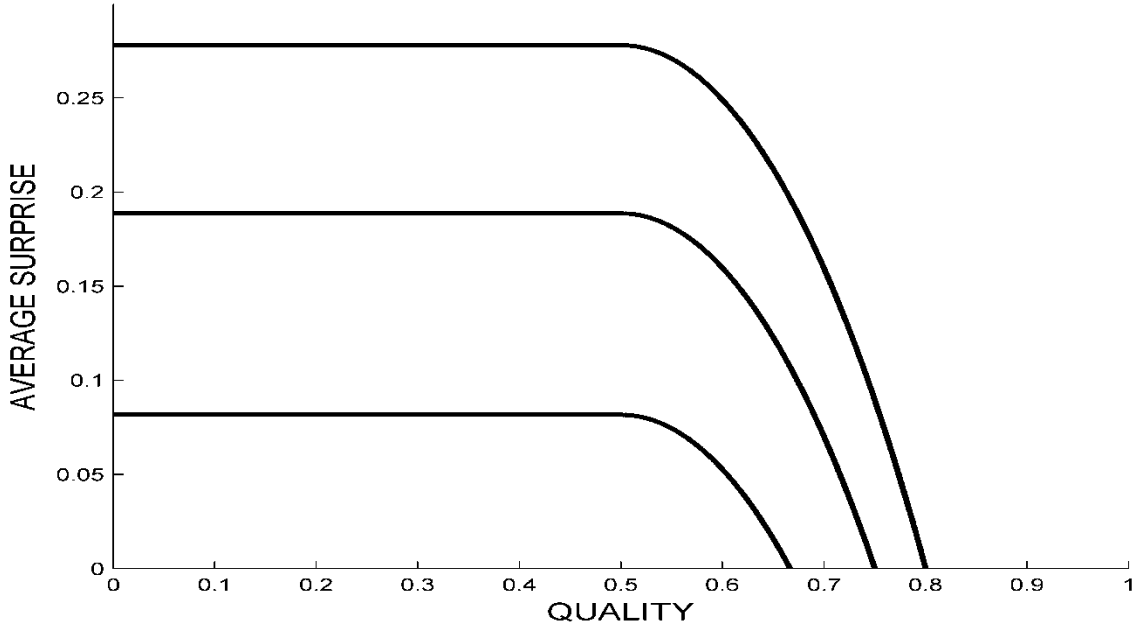


The extremal $P_A(\alpha)$ describes an optimal stochastic sampling algorithm for computational creativity

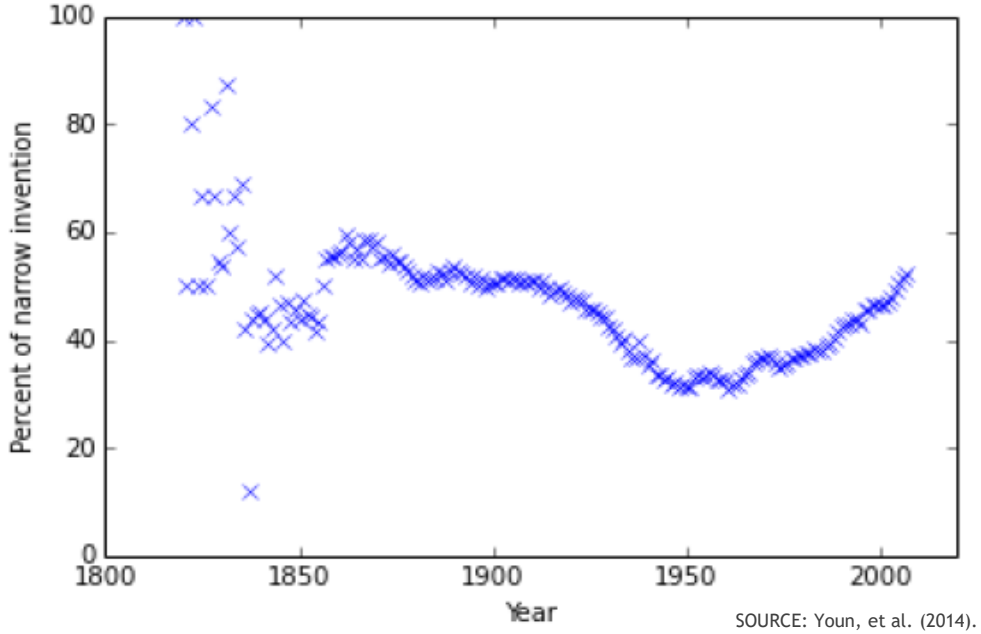
Optimal sequential selection can be analyzed using the theory of concomitants of order statistics

Maturity of the field

- Initially when Θ is very small, $P_{\theta|\alpha}$ may not be absolutely continuous with respect to P_θ, so relative entropy in surprise would be infinite
- After many artifacts are created and known, the effect of the Bayesian belief update due to the new artifact is small
  - Noisier channel shifts curve to left
  - All low-hanging fruits already created



Maturity of the field



SOURCE: Youn, et al. (2014).

- Broad patents were prevalent after WWII, but narrow patents now predominate
  - Growing the component alphabet?
- Time is ripe for broad systems-level inventions, which make use of the fertile resource of narrow inventions

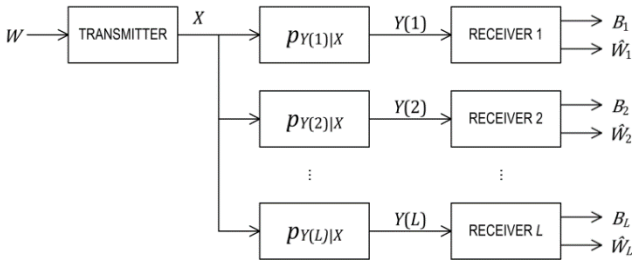
Discovery: Growing the component alphabet

- Lack of absolute continuity in Bayesian surprise expression yields an infinite value
  - Do new components yield transformational creativity that is different in kind from combinatorial creativity?

- Scientific discovery provides new “ingredients” for creating artifacts and ideas, especially if they are high-quality

Diverse audiences

Making the analogy of channel noise with inspiration sets, and of energy functions with quality functions, we obtain the following theorem directly.



Block diagram of multicasting energy and information simultaneously.

(theorem precisely quantifies increased difficulty in creativity when requirements of several audiences must be met simultaneously)

Theorem Consider a creativity setting with several distinct audiences, such that each audience has its own inspiration set Θ_ℓ, quality function q_ℓ, and quality requirement Q_ℓ. Let $\mathfrak{G}_\ell = \{P_A(\alpha)\colon\ \mathbb{E}[q_\ell(A) \geq Q_\ell]\}$ be sets of generative algorithms that meet the ℓth quality constraint and $\mathfrak{G} = \bigcap_{\ell=1}^{L} \mathfrak{G}_\ell$ the set of generative algorithms that meet all quality constraints. Then the *average diverse audience novelty-quality trade-off* is:

$$S(Q_1, \cdots, Q_L) = \max_{P_A(\alpha)\in\mathfrak{G}} \min_{\ell=1,\ldots,L} I(A;\Theta_\ell)\,.$$

Limits of creativity with intentionality

- Consider a communicative intent rate requirement R over a noisy perception channel: $P_{\hat{A}|A}$
- Then the fundamental tradeoff for intentional creativity is as follows

$$S(Q,R) = \max_{P_A(\alpha):E[q(A)]\geq Q,\, I(A,\hat{A})\geq R} I(A,\Theta)$$

- Mathematical theories of creativity
- Generative and creative AI models for art