Capacity Constraints in Cognitive Function

Outline

- Multitasking and Control
- Miller's Law

Outline

☞
- **Multitasking and Control**
- **Miller’s Law**

Cognitive Control

Cognitive Control

This concerns a feature of cognitive control that is its…
simplest...

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

most influential...

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

most influential...

most striking...

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

most influential...

most striking...

and most embarrassing one:

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

most influential...

most striking...

and most embarrassing one:

the *radical* constraint on its engagement

Cognitive Control

This concerns a feature of cognitive control that is its…

simplest...

most basic...

most general...

most influential...

most striking...

and most embarrassing one:

the *radical* constraint on its engagement

There are even *laws* prohibiting allocation of control
to more than one task at a time:

Talking on a cell phone and driving

The Capacity Constraint on Cognitive Control

The Capacity Constraint on Cognitive Control

- **Central to the original theories of cognitive control**
  (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977)

The Capacity Constraint on Cognitive Control

- **Central to the original theories of cognitive control**
  (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977)

- **Continues to define the methodology for studying the engagement of cognitive control: dual task designs**

The Capacity Constraint on Cognitive Control

- **Central to the original theories of cognitive control**
  (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977)

- **Continues to define the methodology for studying the engagement of cognitive control: dual task designs**

- **Foundational assumption of most* current models of cognitive control (e.g., ACT-R, SOAR, PDP models)**

  *though not all (e.g. EPIC)

The Capacity Constraint on Cognitive Control

- **Central to the original theories of cognitive control**
  (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977)

- **Continues to define the methodology for studying the engagement of cognitive control: dual task designs**

- **Foundational assumption of most* current models of cognitive control (e.g., ACT-R, SOAR, PDP models)**
  *though not all (e.g. EPIC)

- **Yet, an *explanation* for this constraint remains elusive**

The Capacity Constraint on Cognitive Control

- **Central to the original theories of cognitive control**
  (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977)

- **Continues to define the methodology for studying the engagement of cognitive control: dual task designs**

- **Foundational assumption of most* current models of cognitive control (e.g., ACT-R, SOAR, PDP models)**
  *though not all (e.g. EPIC)

- **Yet, an *explanation* for this constraint remains elusive**

What is its *cause,* and *why* is it so *restrictive?*

Capacity Constraints

Capacity Constraints

- “Explanations” commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of “slots”)

Capacity Constraints

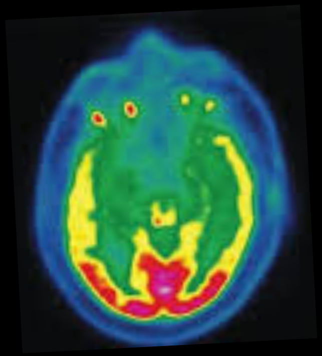
- "Explanations" commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of "slots")
- Those explanations are silly

Capacity Constraints

- “Explanations” commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of “slots”)
- Those explanations are silly
  - Metabolic?

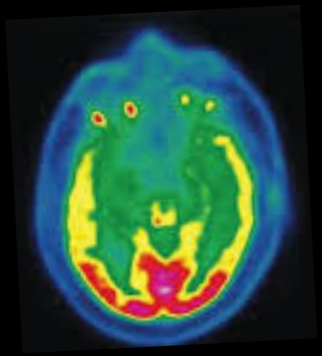
Capacity Constraints

- "Explanations" commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of "slots")
- Those explanations are silly
  - Metabolic?
    - Vast areas of cortex (e.g., visual) are "running" all the time



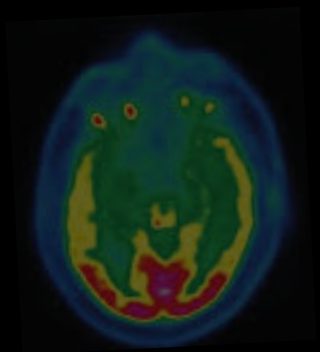
Capacity Constraints

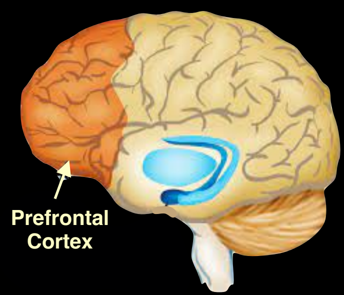
- "Explanations" commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of "slots")
- Those explanations are silly
  - Metabolic?
    - Vast areas of cortex (e.g., visual) are "running" all the time
  - Structural?



Capacity Constraints

- "Explanations" commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of "slots")
- Those explanations are silly
  - Metabolic?
    - Vast areas of cortex (e.g., visual) are "running" all the time
  - Structural?
    - PFC comprises ~30% of neocortex = ~30 billion neurons... but only 1 control-demanding task at a time? Really?

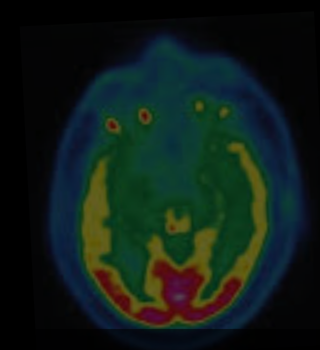


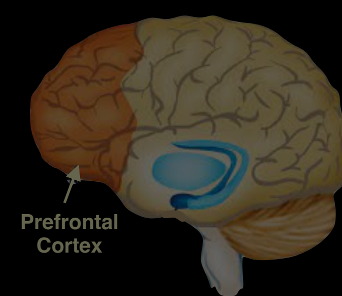


Capacity Constraints

- "Explanations" commonly offered:
  - metabolic (energetically costly) or
  - structural (limited number of "slots")

- Those explanations are silly
  - Metabolic?
    - Vast areas of cortex (e.g., visual) are "running" all the time
  - Structural?
    - PFC comprises ~30% of neocortex = ~30 billion neurons… but only 1 control-demanding task at a time? Really?





- **The question begs a more rational explanation…**

Capacity Constraints

Capacity Constraints

- **Structural models:**
  - **constraints of attractor systems**
    (Usher et al., 2001)

Capacity Constraints

- **Structural models:**
  - **constraints of attractor systems**
    (Usher et al., 2001)
  - **capacity vs. precision**
    (e.g., Ma & Huang, 2009; Luck & Vogel, 2013, etc.)

Capacity Constraints

- **Structural models:**
  - **constraints of attractor systems**
    (Usher et al., 2001)
  - **capacity vs. precision**
    (e.g., Ma & Huang, 2009; Luck & Vogel, 2013, etc.)
    - **These accounts still beg the question of *why so strict,* given ~30B neurons?**

Capacity Constraints

- **Structural models:**
  - **constraints of attractor systems**
    (Usher et al., 2001)
  - **capacity vs. precision**
    (e.g., Ma & Huang, 2009; Luck & Vogel, 2013, etc.)
    - **These accounts still beg the question of *why so strict,* given ~30B neurons?**

- **Functional (computational) accounts:**
  - **Symbolic models (production system architectures)**
    (e.g., Meyer & Kieras, 1997; Salvucci & Taatgren, 2008)
    - **Scheduling constraints to avoid cross-talk imposed by interacting processes**
    - **Assume pre-specified processing architecture**

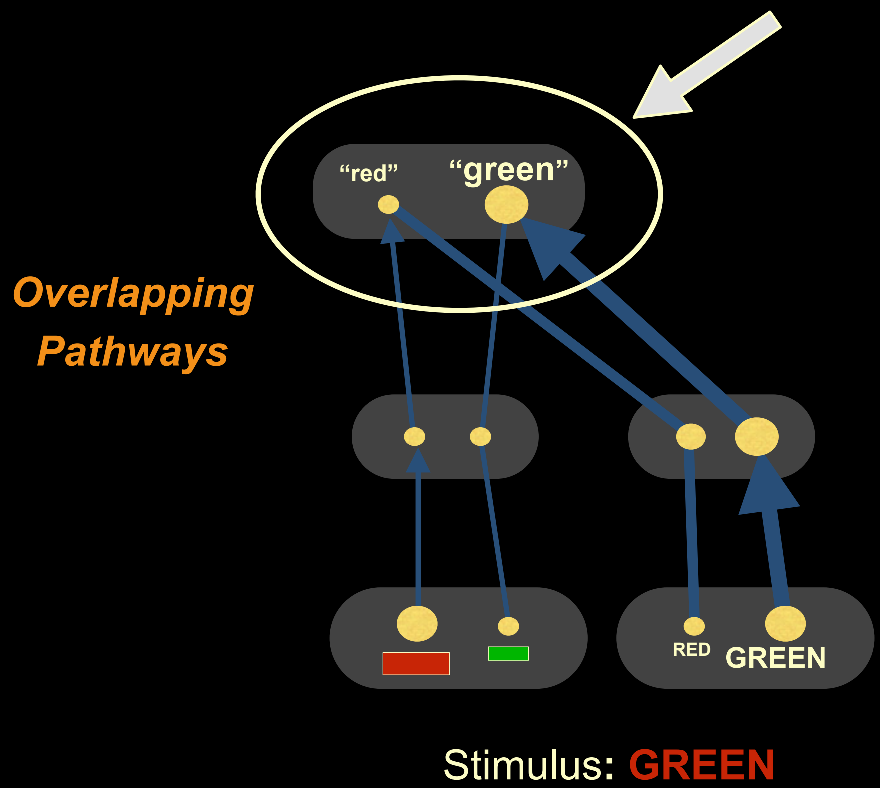
Capacity Constraints

- **Structural models:**
  - **constraints of attractor systems**
    (Usher et al., 2001)
  - **capacity vs. precision**
    (e.g., Ma & Huang, 2009; Luck & Vogel, 2013, etc.)
    - **These accounts still beg the question of *why so strict,* given ~30B neurons?**

- **Functional (computational) accounts:**
  - **Symbolic models (production system architectures)**
    (e.g., Meyer & Kieras, 1997; Salvucci & Taatgren, 2008)
    - **Scheduling constraints to avoid cross-talk imposed by interacting processes**
    - **Assume pre-specified processing architecture**
    - **Beg the question: *why that architecture?***

Shared Representation vs. Multitasking

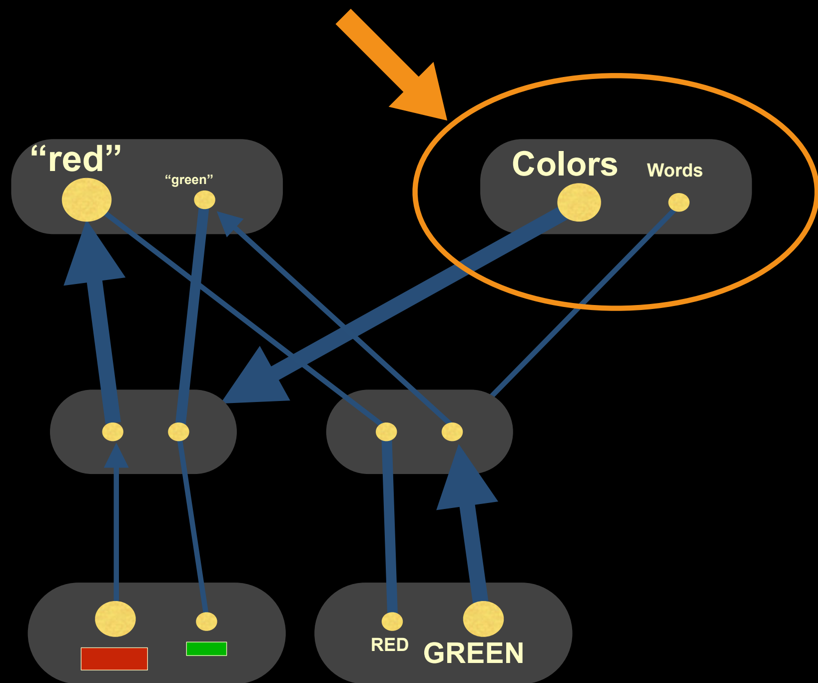
- **Basic idea:**
  - **Pathway overlap (shared representations) introduces potential for *cross-talk***



Shared Representation vs. Multitasking

- **Basic idea:**
  - Pathway overlap (shared representations) introduces potential for cross-talk
  - **The purpose of control is to manage this**

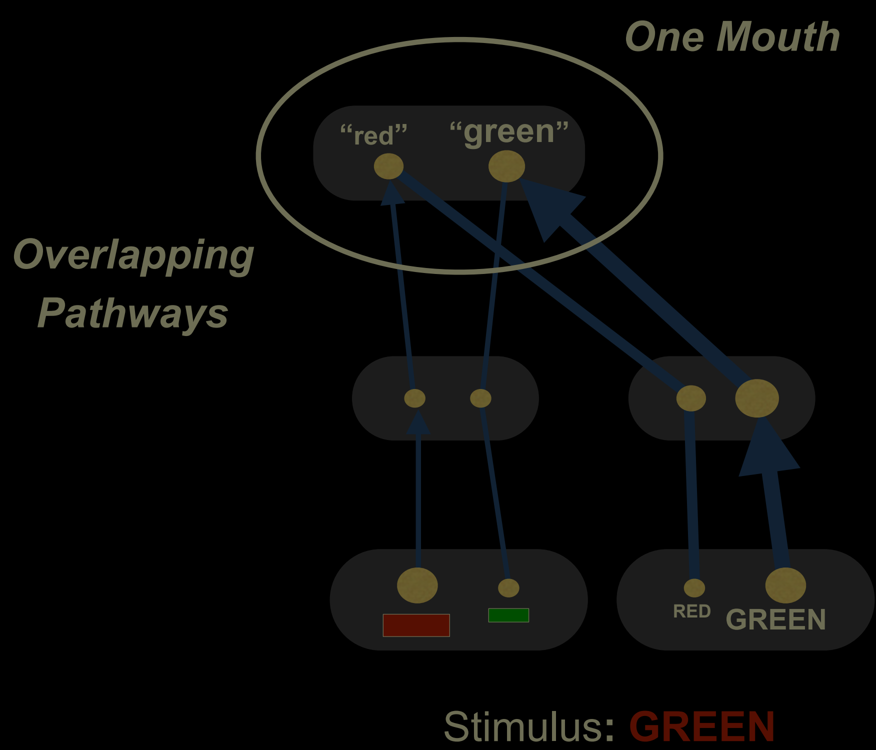
Overlapping Pathways



Stimulus: GREEN

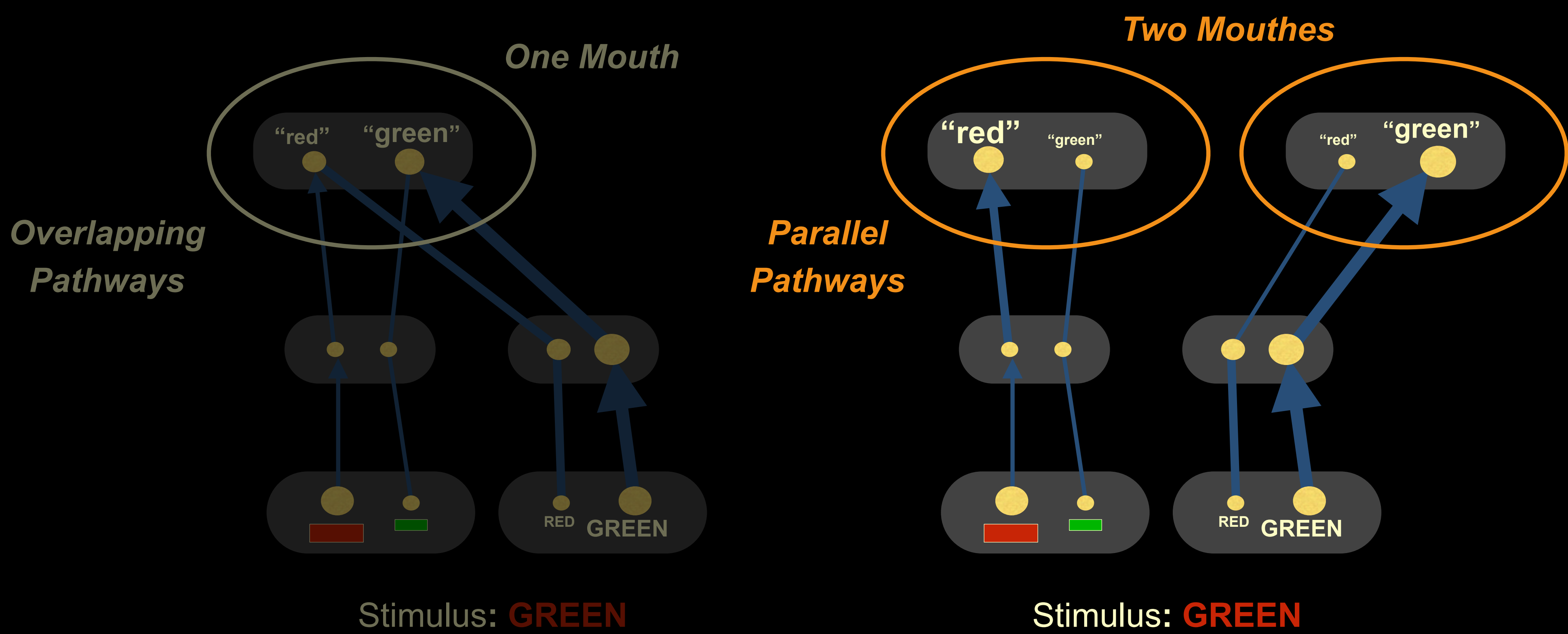
Shared Representation vs. Multitasking

- **Basic idea:**
  - Pathway overlap (shared representations) introduces potential for cross-talk
  - The purpose of control is to manage this
  - Can solve the problem without control…



Shared Representation vs. Multitasking

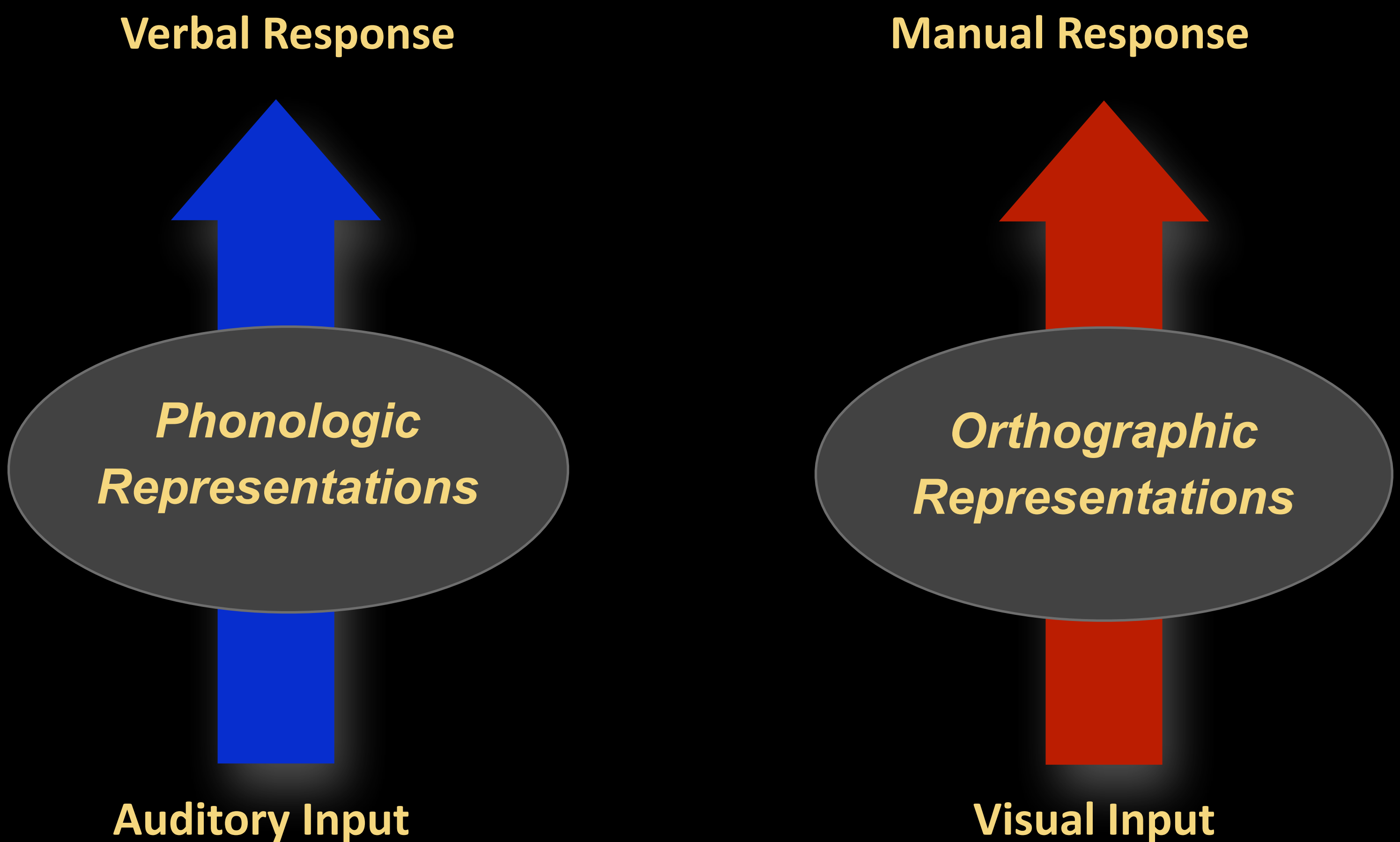
- **Basic idea:**
  - Pathway overlap (shared representations) introduces potential for cross-talk
  - The purpose of control is to manage this
  - Can solve the problem without control… but that carries its own costs:



Shared Representation vs. Multitasking

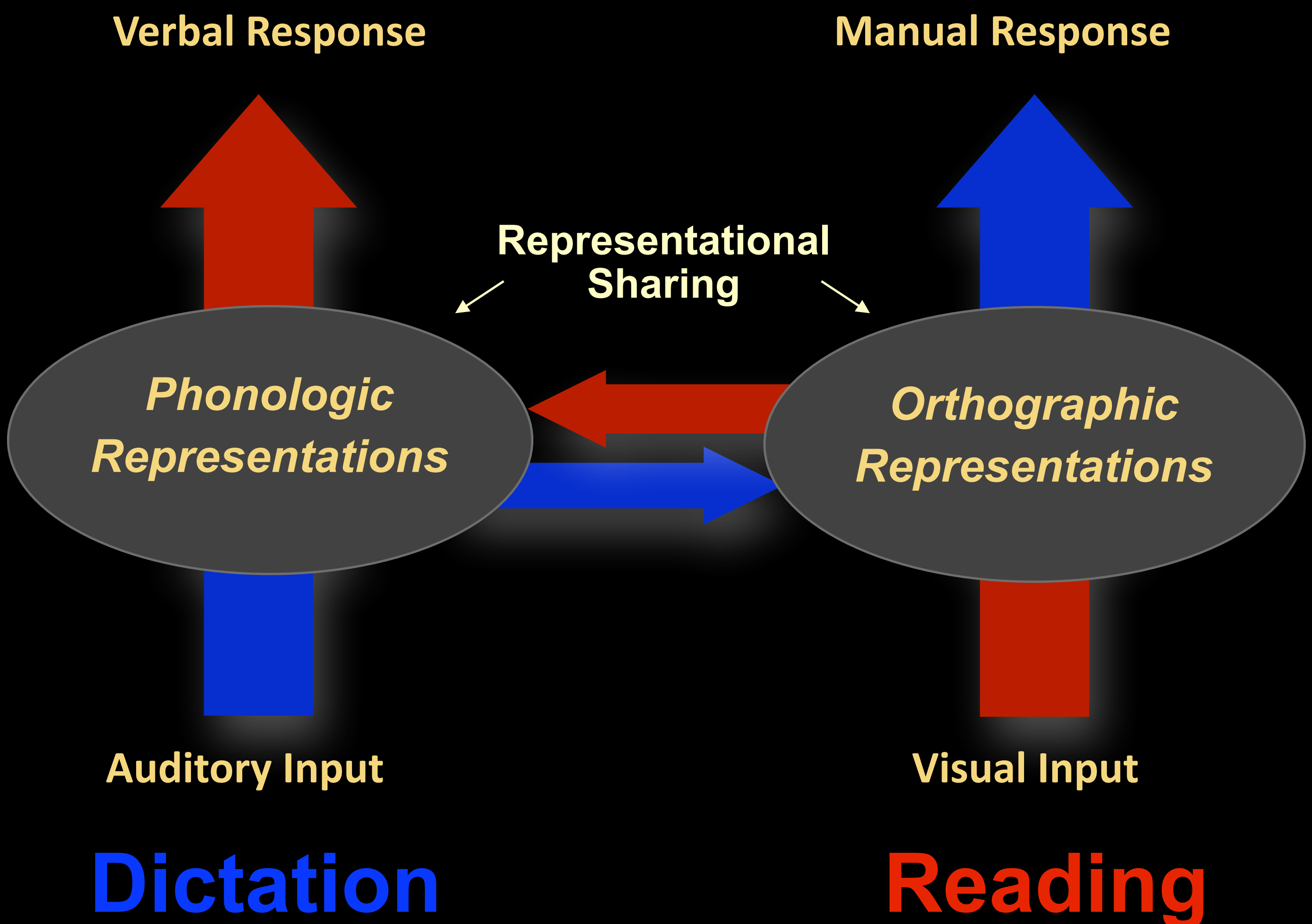
- **Classic illustration:**
  (Shaffer et al., 1975)
  - ***echoing*** **a speech stream while** ***copy-typing*** **(easy) vs.**
  - ***dictation*** **while** ***reading aloud*** **(hard)**

Shared Representation vs. Multitasking



Echoing **Copy-Typing**

Shared Representation vs. Multitasking



Crossings

Crossings

Crossings

Crossings

Rational Boundedness of Control

Rational Boundedness of Control

- Constraints on control-dependent processing reflect:
  - the *representational cost* of shared representation and the role of control in avoiding this cost

Rational Boundedness of Control

- Constraints on control-dependent processing reflect:
  - the *representational cost* of shared representation and the role of control in avoiding this cost
  - *not* limitations in the control mechanism *itself*

Rational Boundedness of Control

- Constraints on control-dependent processing reflect:
  - the *representational cost* of shared representation and the role of control in avoiding this cost
  - *not* limitations in the control mechanism *itself*

- Inverts the standard interpretation:

Control reflects ***bound on rational processing***

Rational Boundedness of Control

- Constraints on control-dependent processing reflect:
  - the *representational cost* of shared representation and the role of control in avoiding this cost
  - *not* limitations in the control mechanism *itself*

- Inverts the standard interpretation:

Control reflects ***rational bound on*** ***processing***

Shared Representation vs. Multitasking

Shared Representation vs. Multitasking

- **Descendant of the Multiple Resources theory of attention**
  Navon & Gopher (1979); Allport (1982); Meyer & Kieras, (1997)
  - *B***ut that was a *qualitative* theory; also, what are *resources?***

Shared Representation vs. Multitasking

- **Descendant of the Multiple Resources theory of attention**
  Navon & Gopher (1979); Allport (1982); Meyer & Kieras, (1997)
  - *B*ut that was a ***qualitative*** theory; also, what are ***resources?***

- **Question:**
  - **How does the maximum number of processes (multitasking) *scale* with potential crosstalk (shared representations) in networks?**

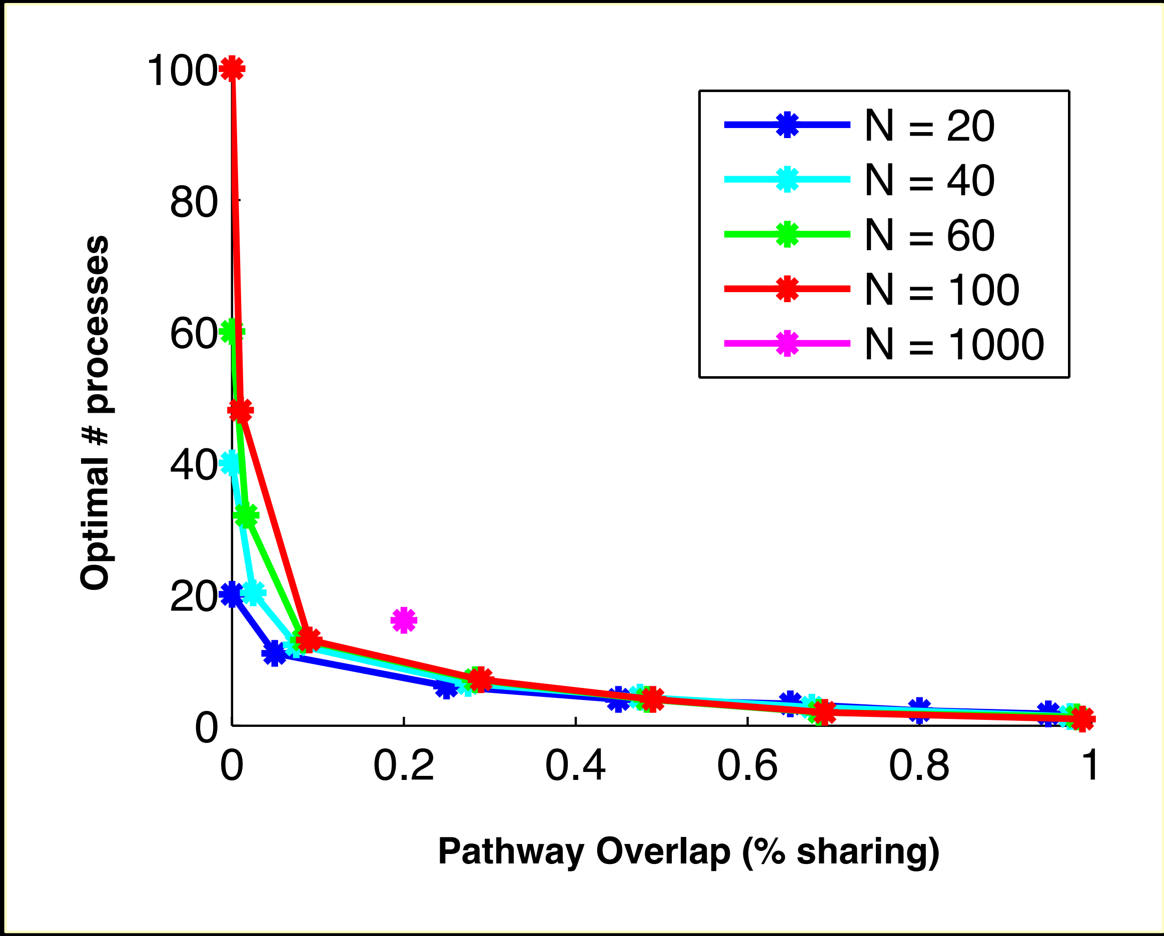
Shared Representation vs. Multitasking

- **Descendant of the Multiple Resources theory of attention**
  Navon & Gopher (1979); Allport (1982); Meyer & Kieras, (1997)
  - *B*ut that was a ***qualitative*** theory; also, what are ***resources?***
- **Question:**
  - **How does the maximum number of processes (multitasking) *scale* with potential crosstalk (shared representations) in networks?**
- **Answer:**
  - **It scales *badly!***
    - Modest amounts of overlap (sharing) *dramatically* limits multitasking
    - Constraints on multitasking reflect this tradeoff

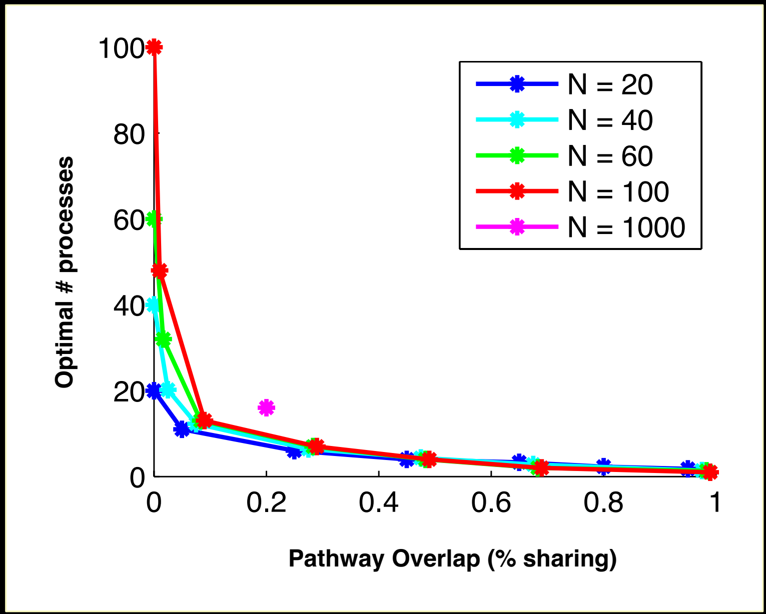
Shared Representation vs. Multitasking

- **Descendant of the Multiple Resources theory of attention**
  Navon & Gopher (1979); Allport (1982); Meyer & Kieras, (1997)
  - *B*ut that was a ***qualitative*** theory; also, what are ***resources?***
- **Question:**
  - **How does the maximum number of processes (multitasking) *scale* with potential crosstalk (shared representations) in networks?**
- **Answer:**
  - **It scales *badly!***
    - Modest amounts of overlap (sharing) *dramatically* limits multitasking
    - Constraints on multitasking reflect this tradeoff
- **Formal treatment:**
  - **Numerical analysis** (simulations in a variety of network architectures)
  - **Mathematical analysis** (graph theoretic analyses)

Pathway Overlap (Shared Representation)



Pathway Overlap (Shared Representation)



Why Shared Representations?

Why Shared Representations?

- Learning benefits by shared representations

Why Shared Representations?

- **But this costs us in the ability to multitask** *(i.e., do two control-demanding tasks at once)*

Why Shared Representations?

- To see this, let's do a little demonstration…

Color Naming + Location Pointing

Name the <u>color</u> of the stimulus and
at the same time point to <u>where</u> it is…

BLUE

YELLOW

Word Mapping

Point **left** if the written word is

⬅ RED

Point **right** if the written word is

GREEN ➡

GREEN

RED

Color Naming + Word Mapping

Name the color of the following stimulus
and, *at the same time*:

Point **left** if the written word is

⬅ RED

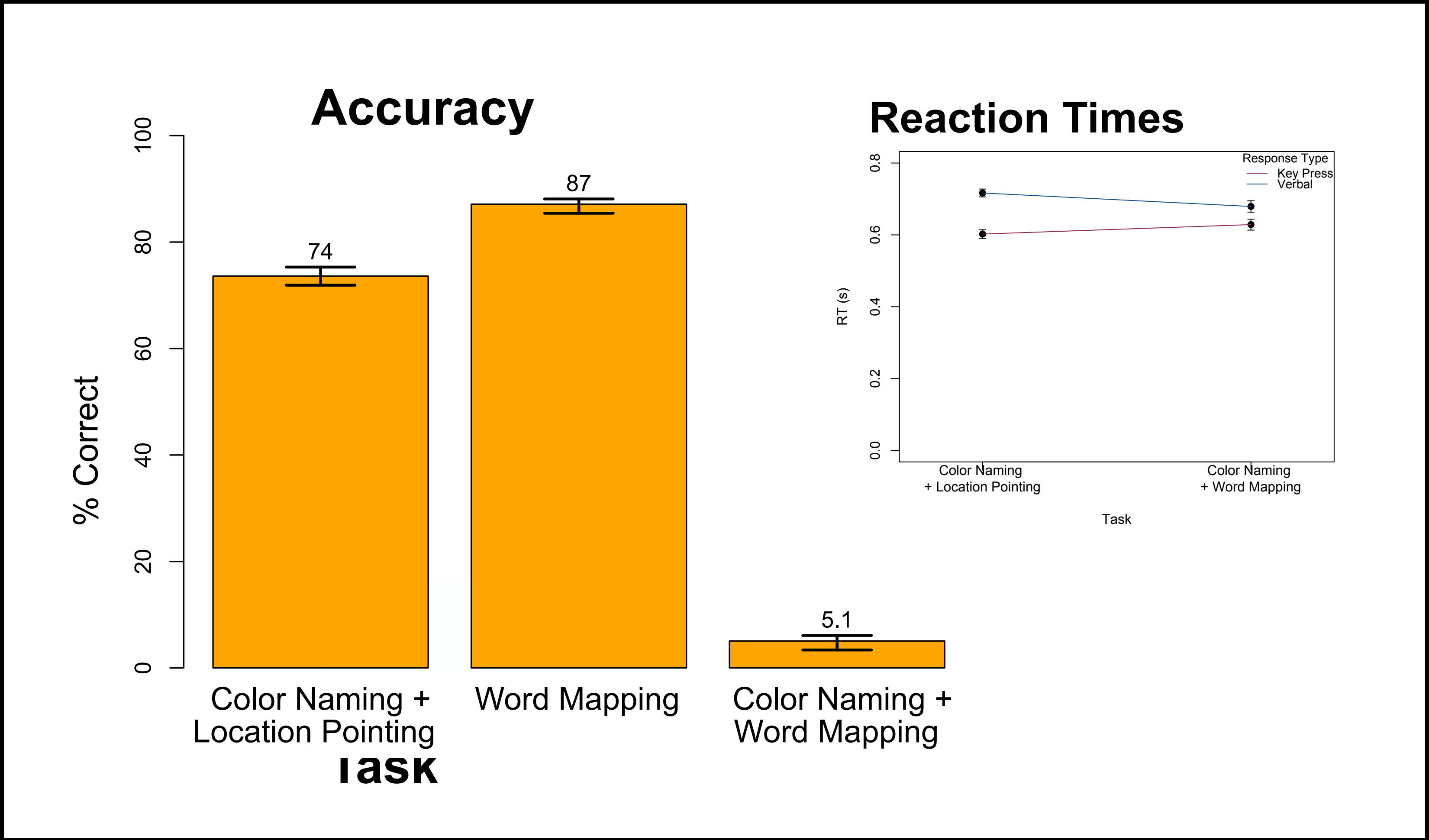
Point **right** if the written word is

GREEN ➡

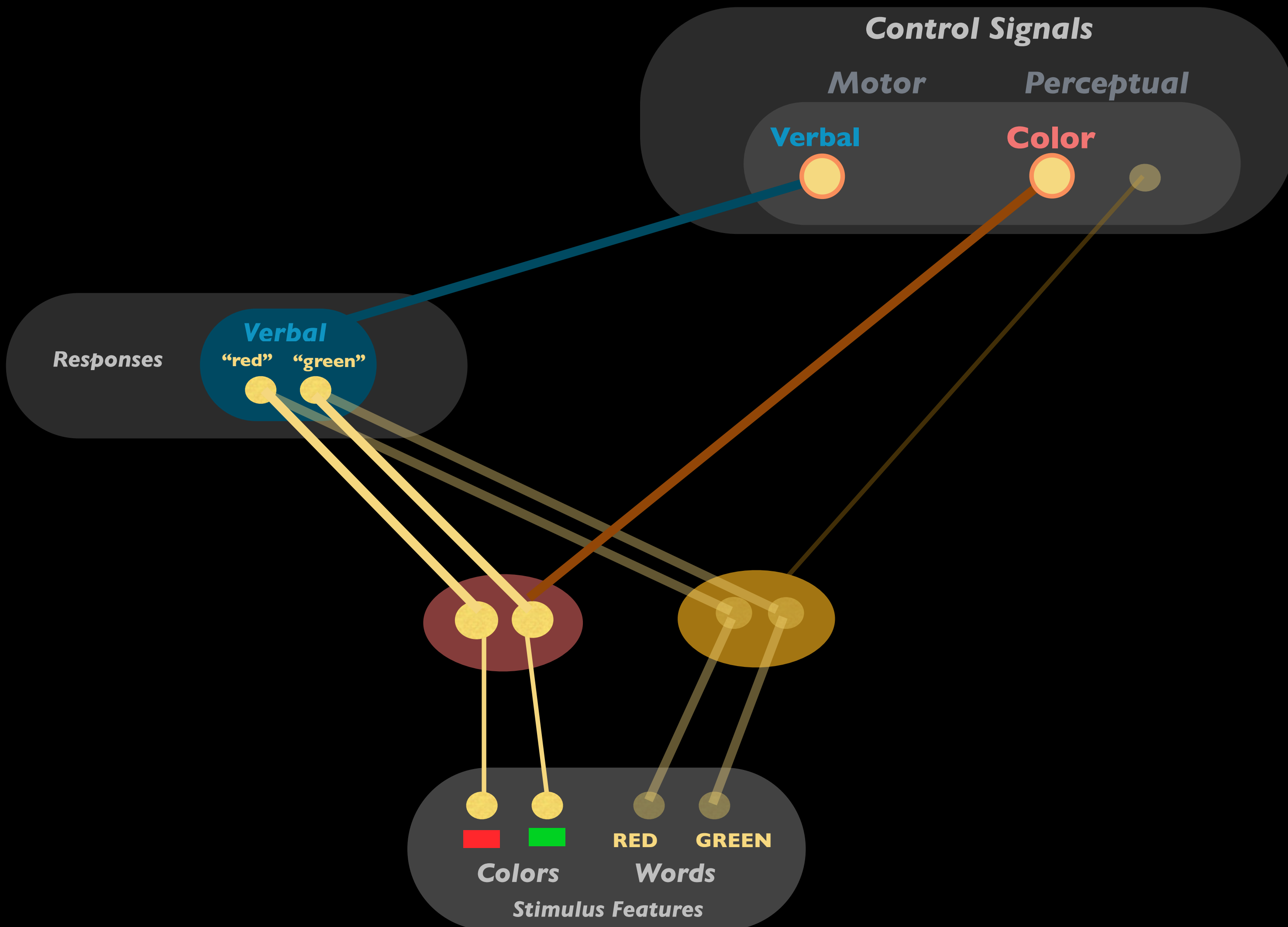
GREEN

RED

Learning and Multitasking

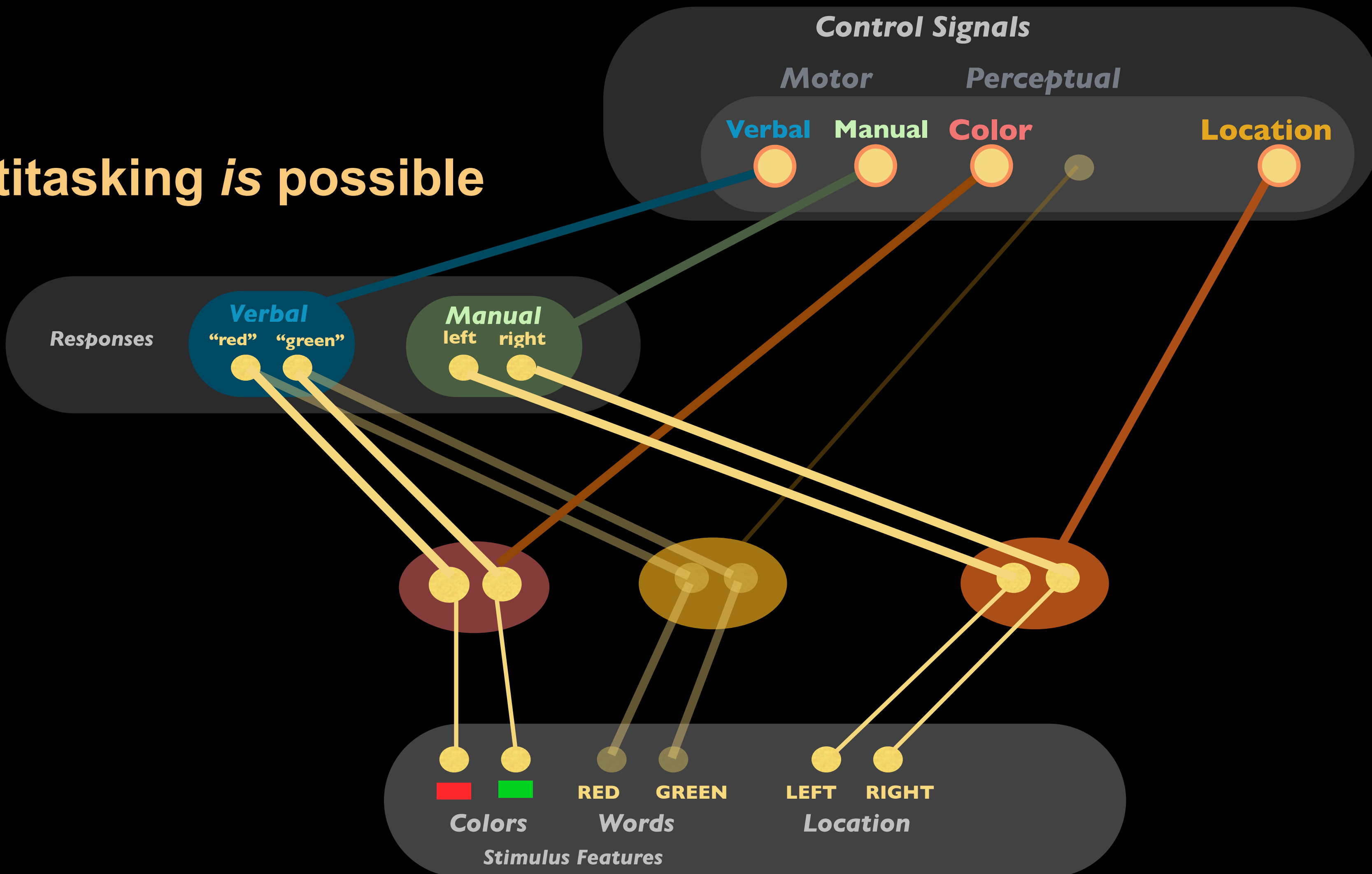


Color Naming



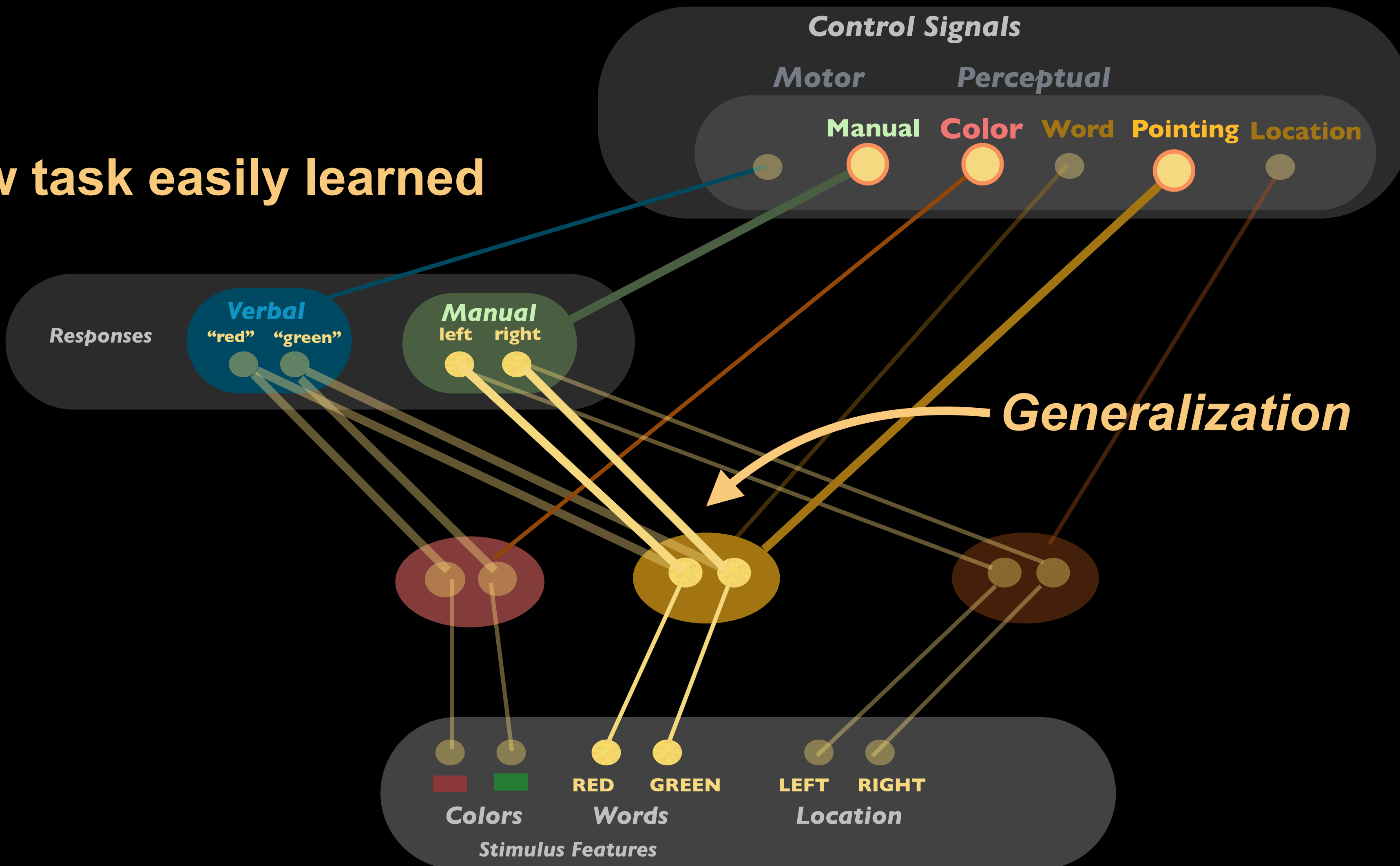
Color Naming + Location Pointing

Multitasking *is* possible



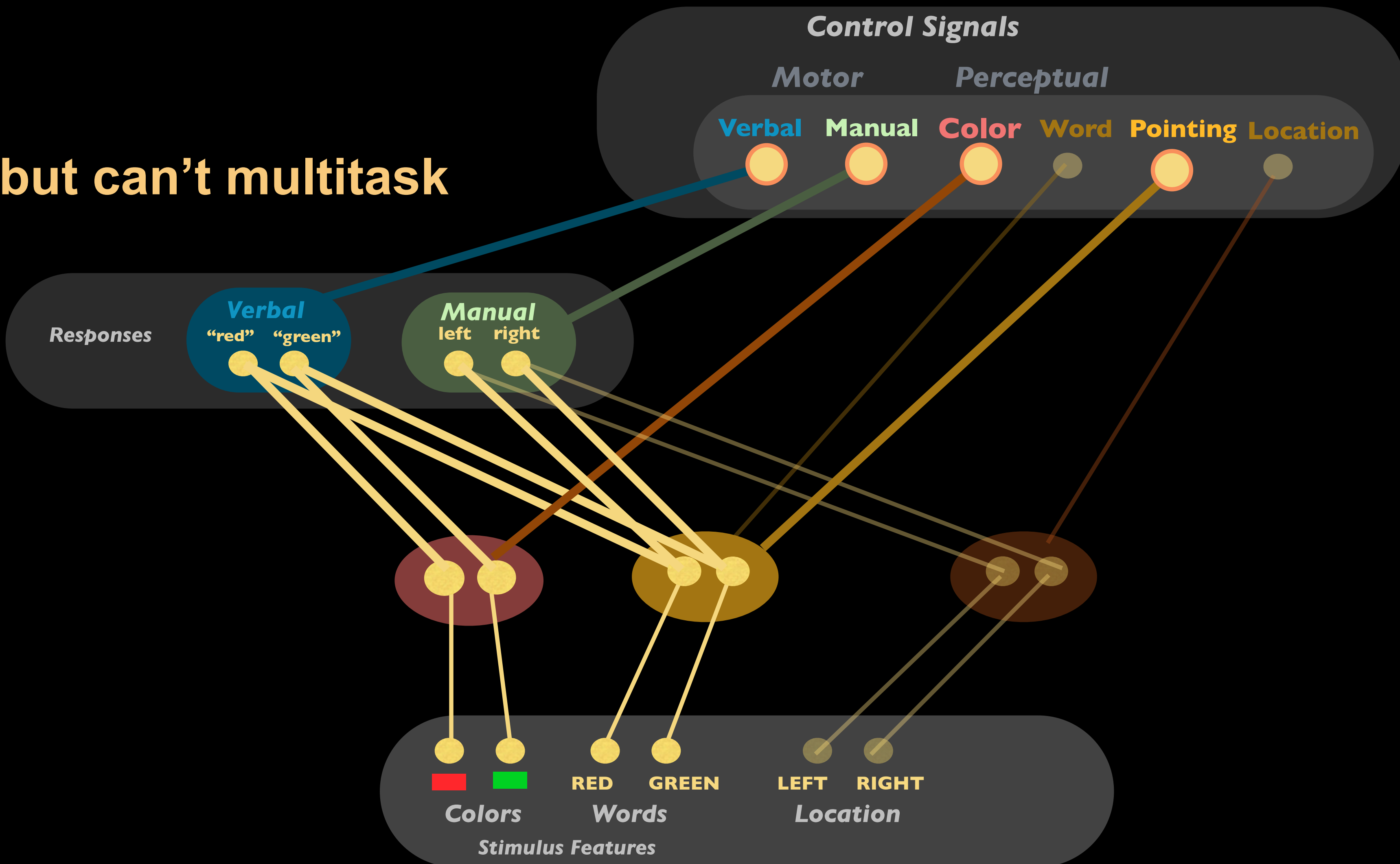
Word Mapping

New task easily learned



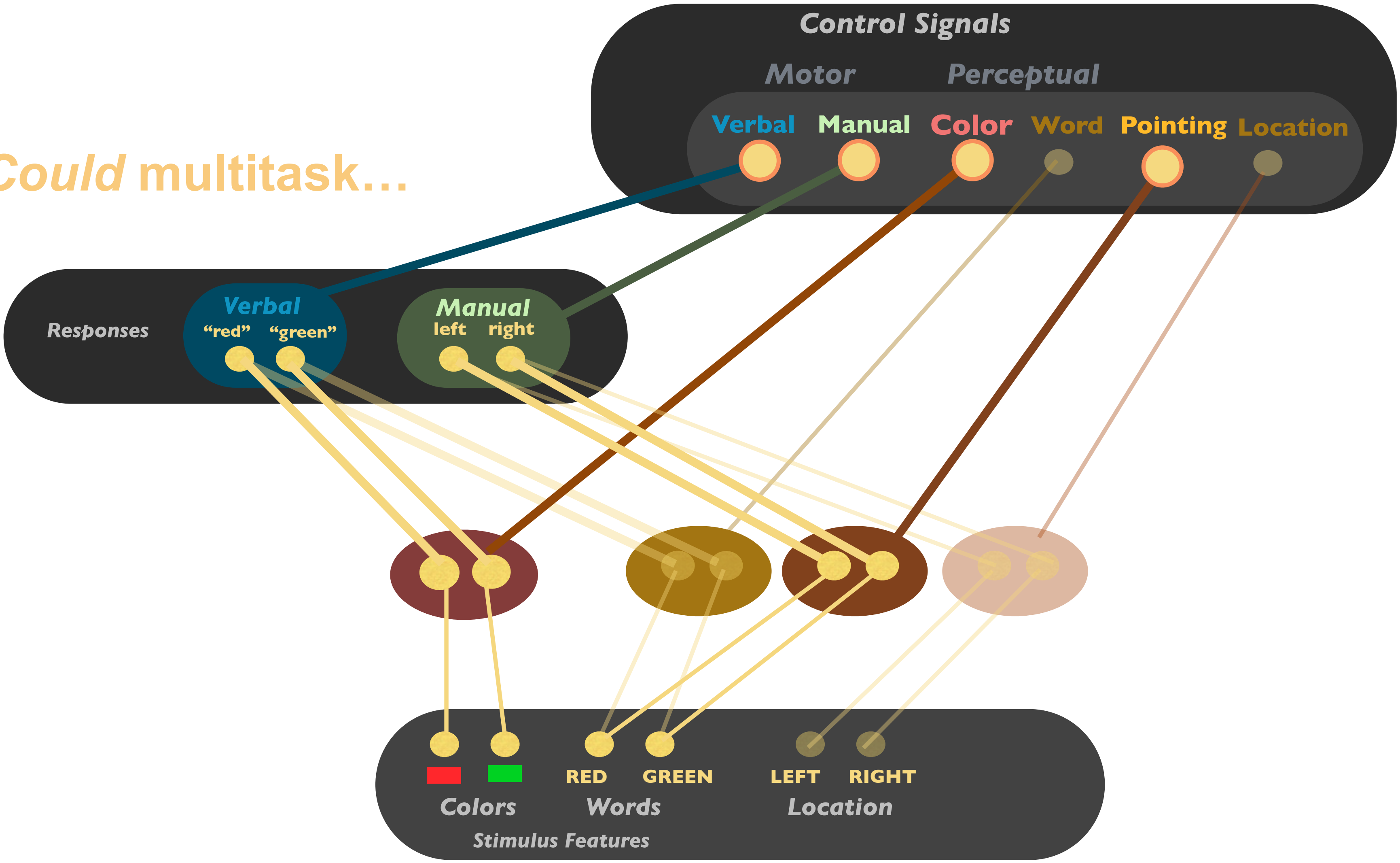
Color Naming + Word Mapping

…but can't multitask



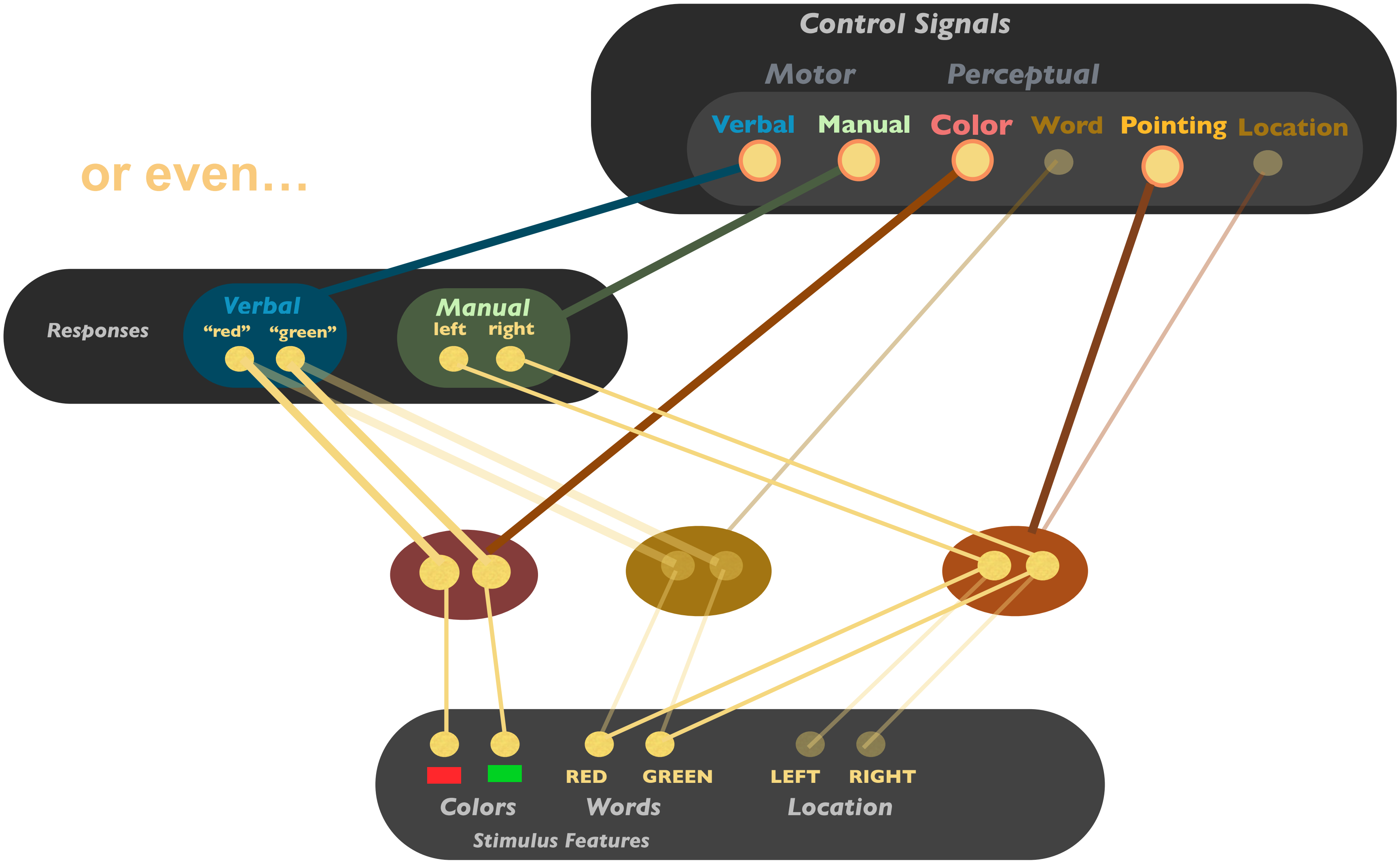
Color Naming + Word Mapping

Could multitask…

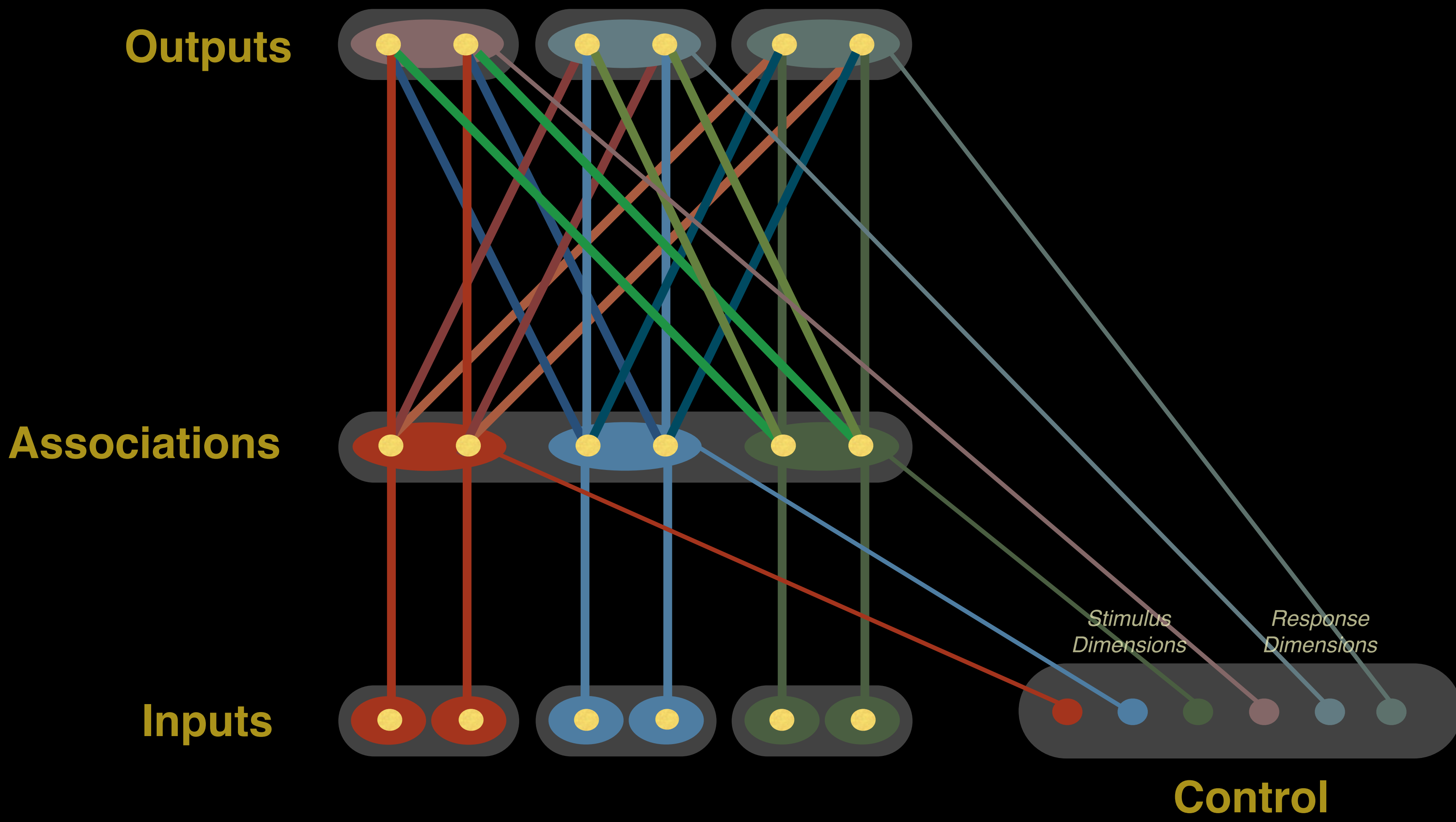


Color Naming + Word Mapping

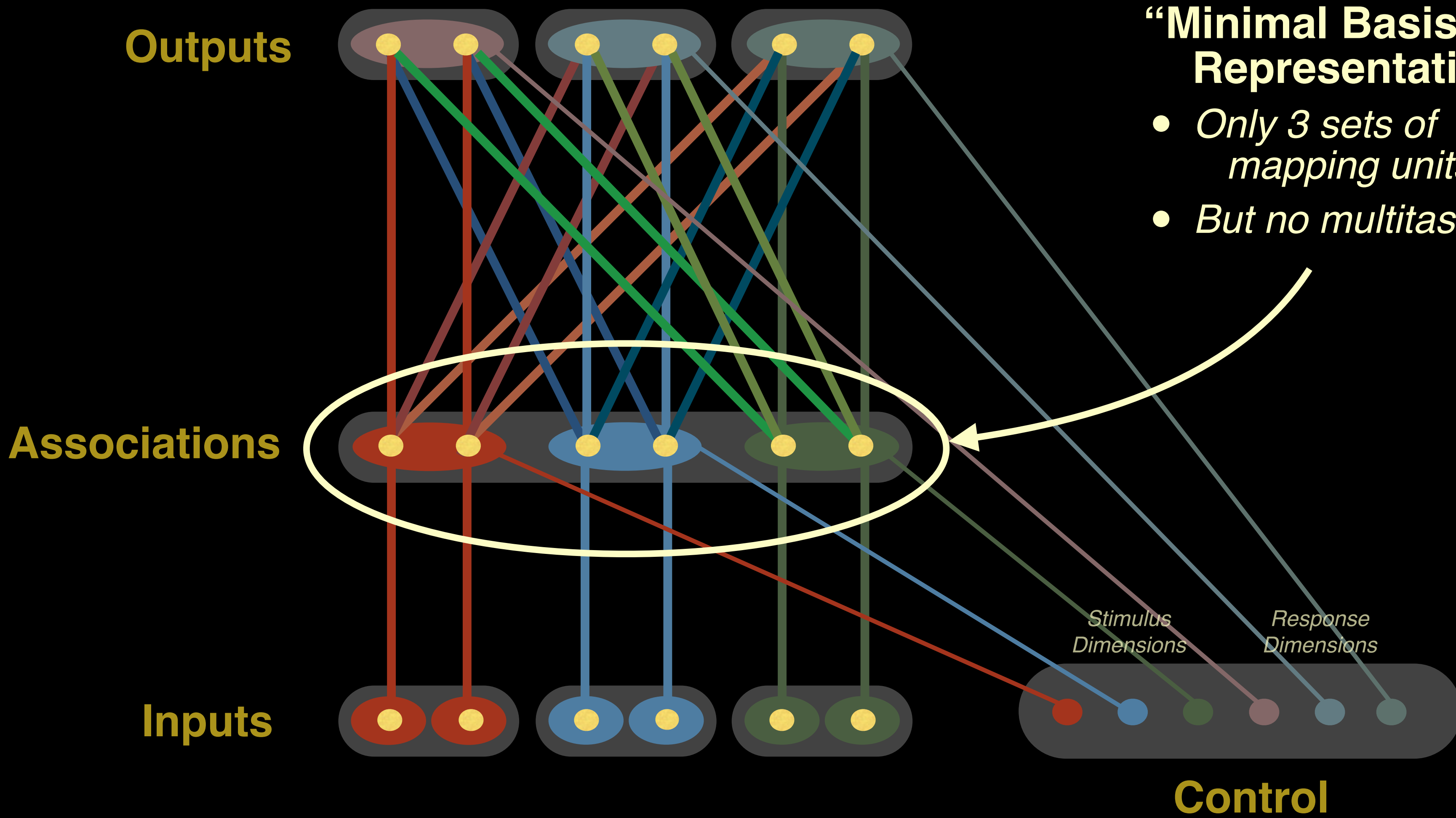
or even…



Network Representation of Multiple Tasks



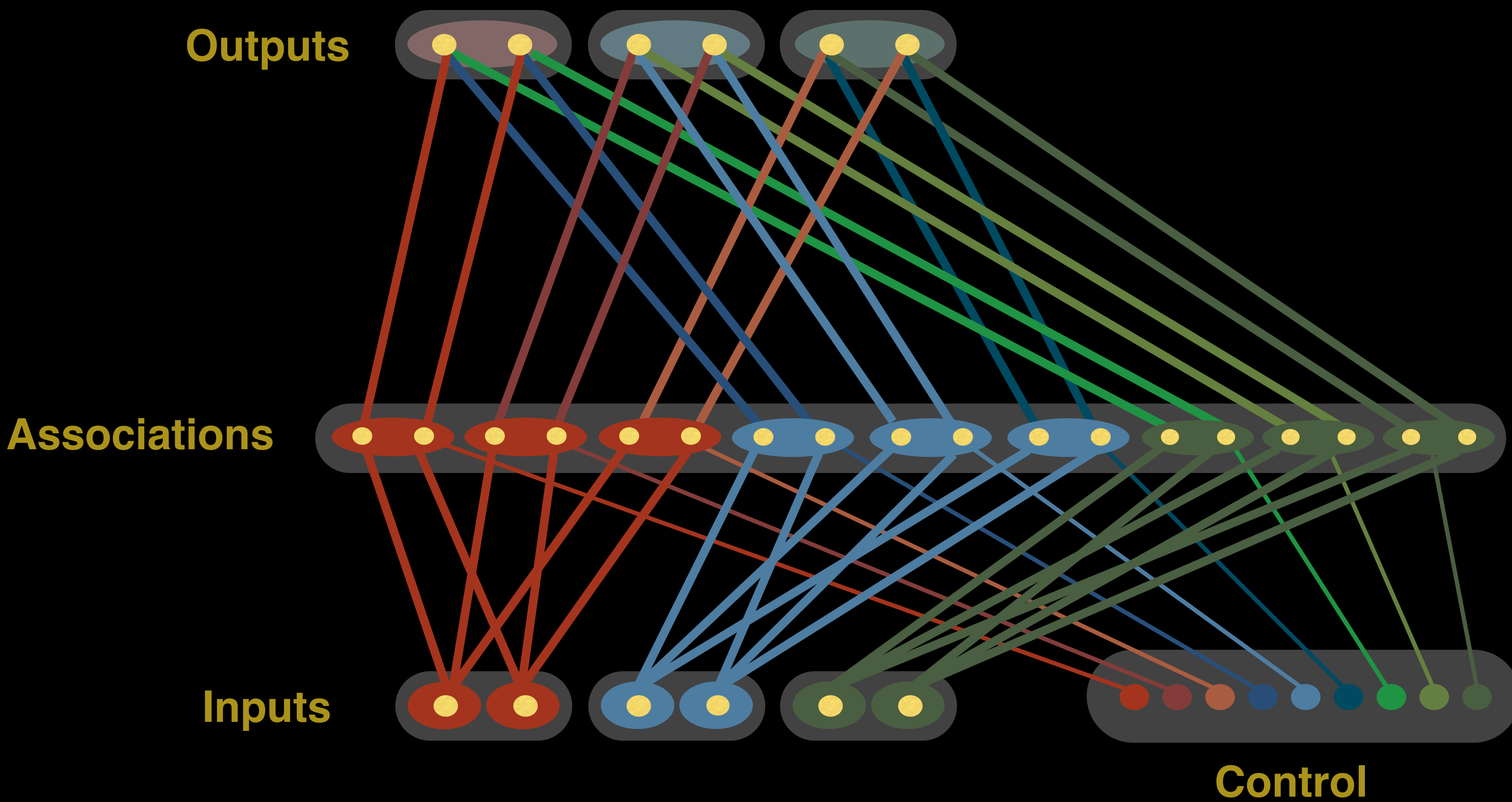
Network Representation of Multiple Tasks



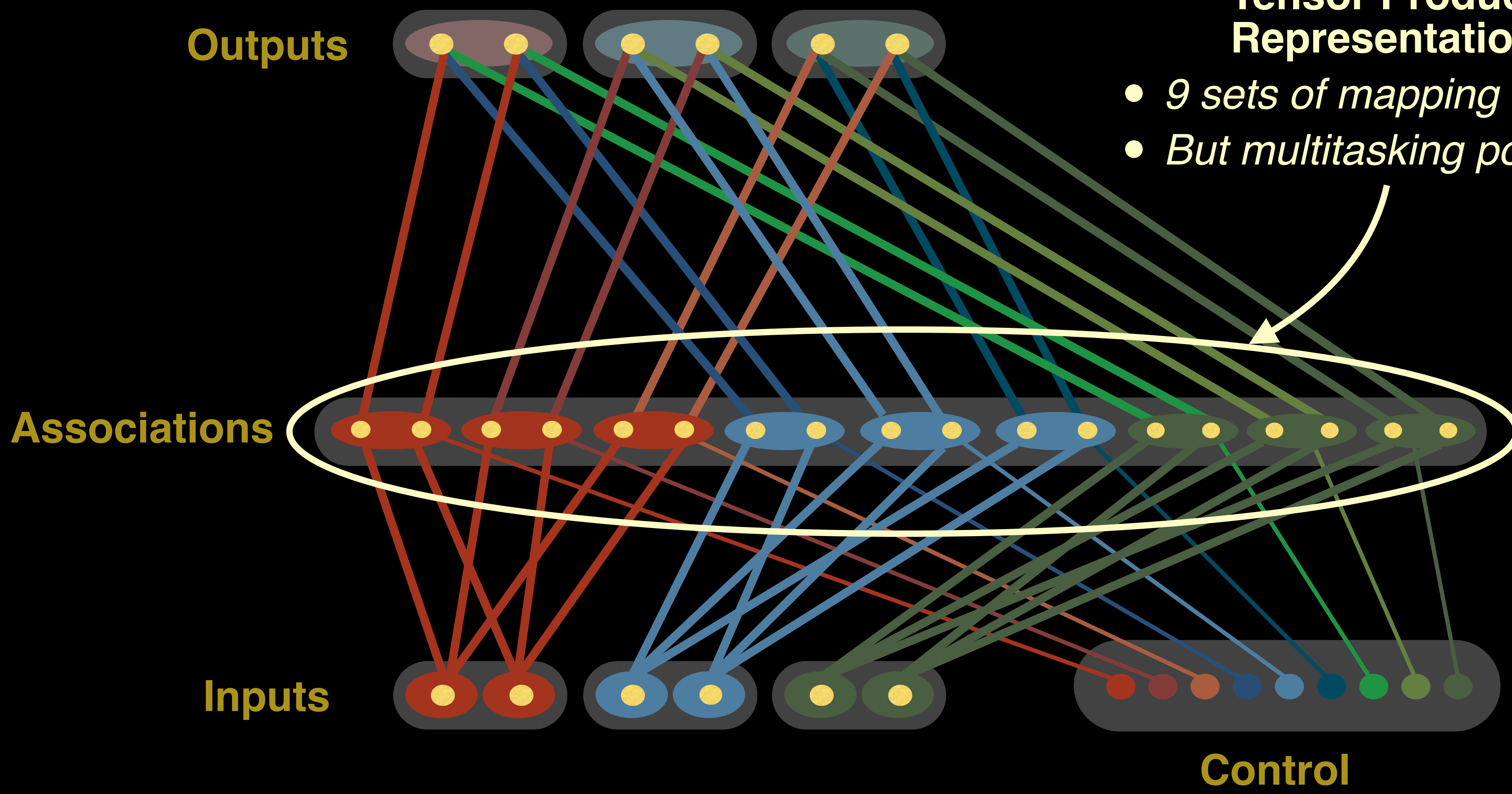
"Minimal Basis Set" Representation

- *Only 3 sets of mapping units*
- *But no multitasking*

Network Representation of Multiple Tasks



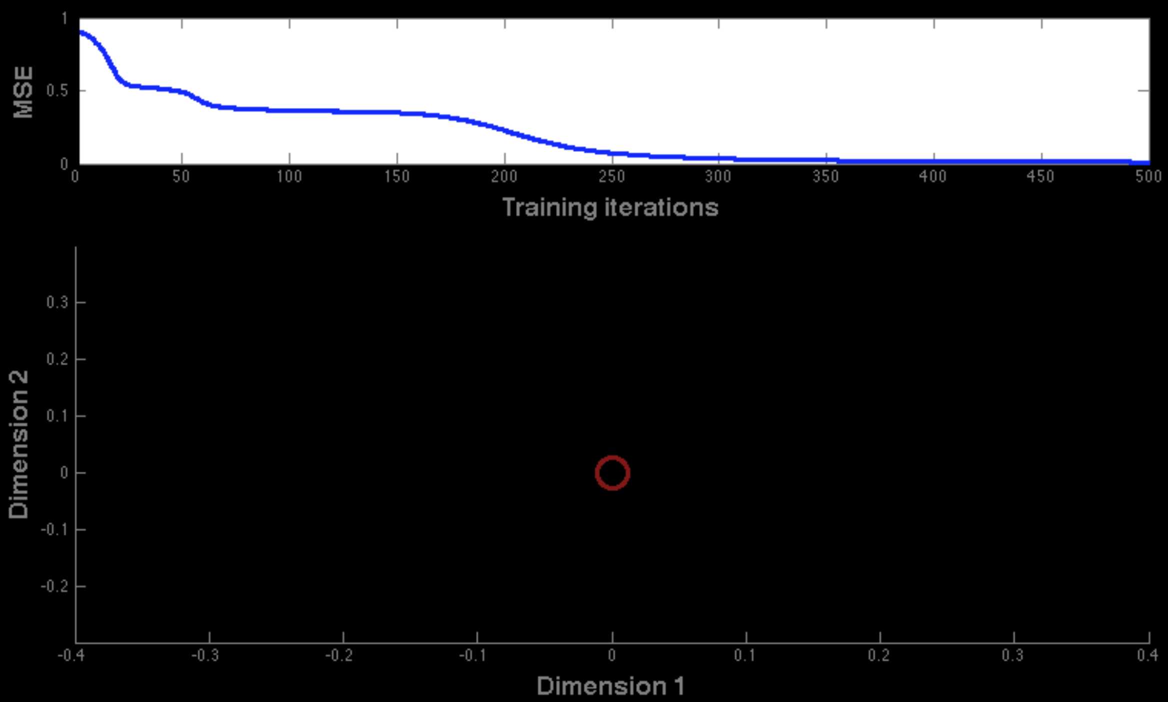
Network Representation of Multiple Tasks



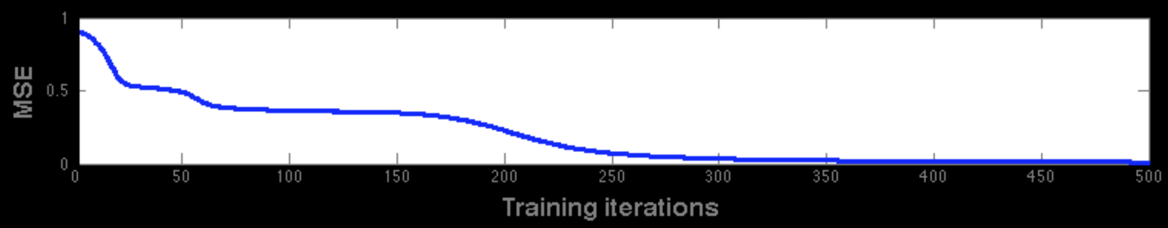
"Tensor Product" Representation

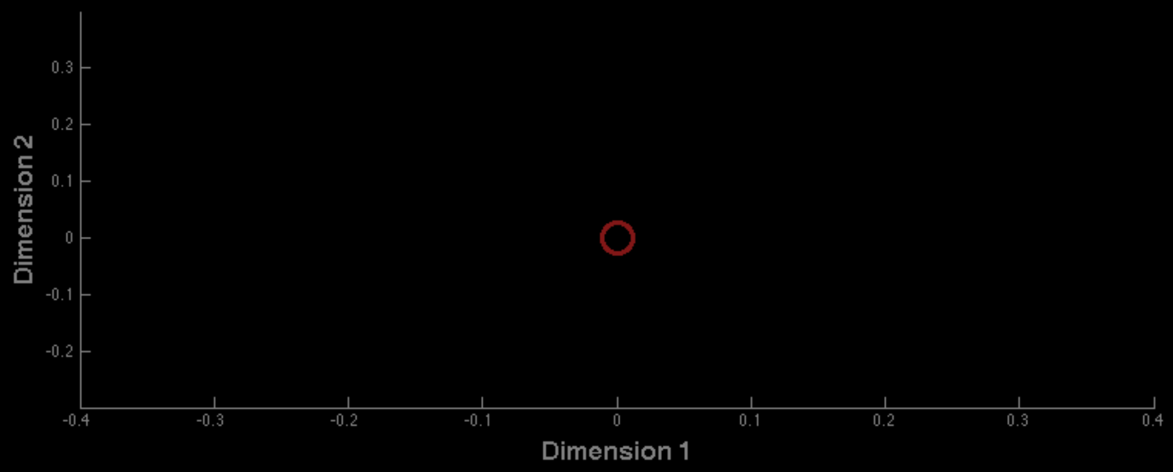
- *9 sets of mapping units*
- *But multitasking possible*

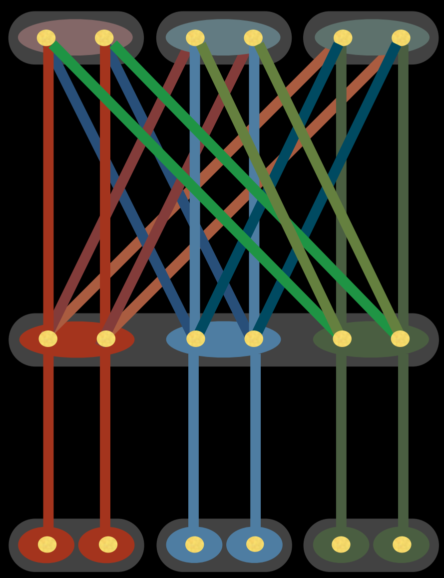
Single Task Training



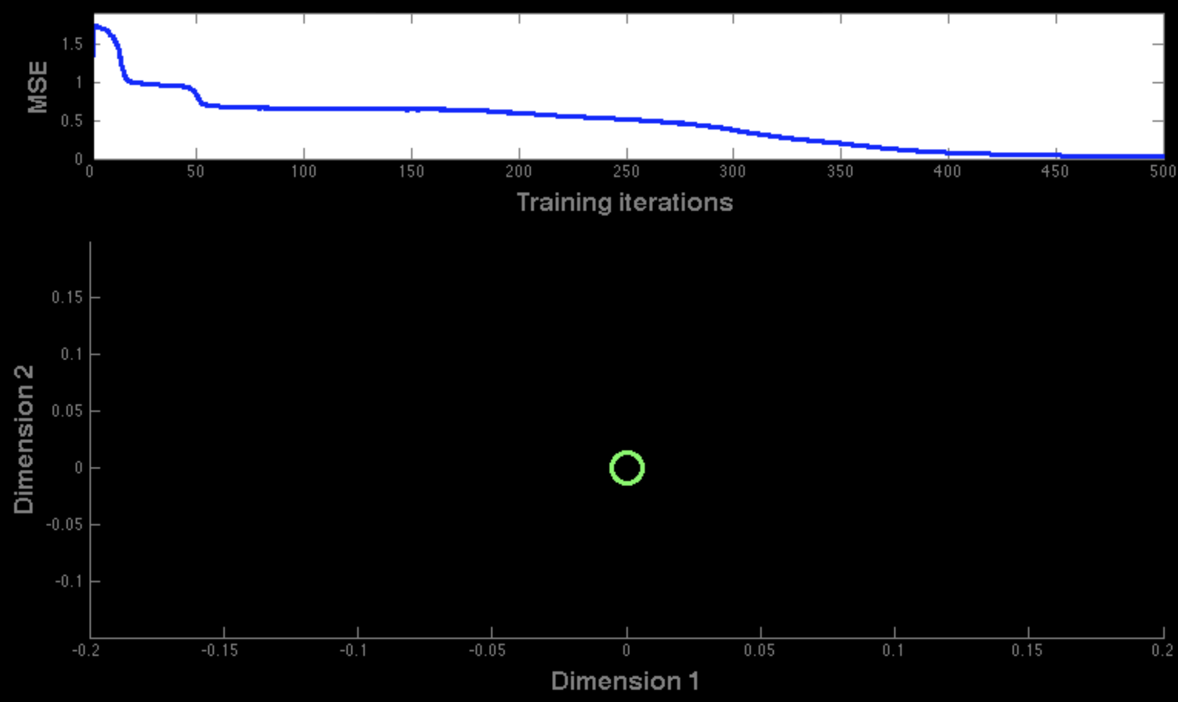
Single Task Training



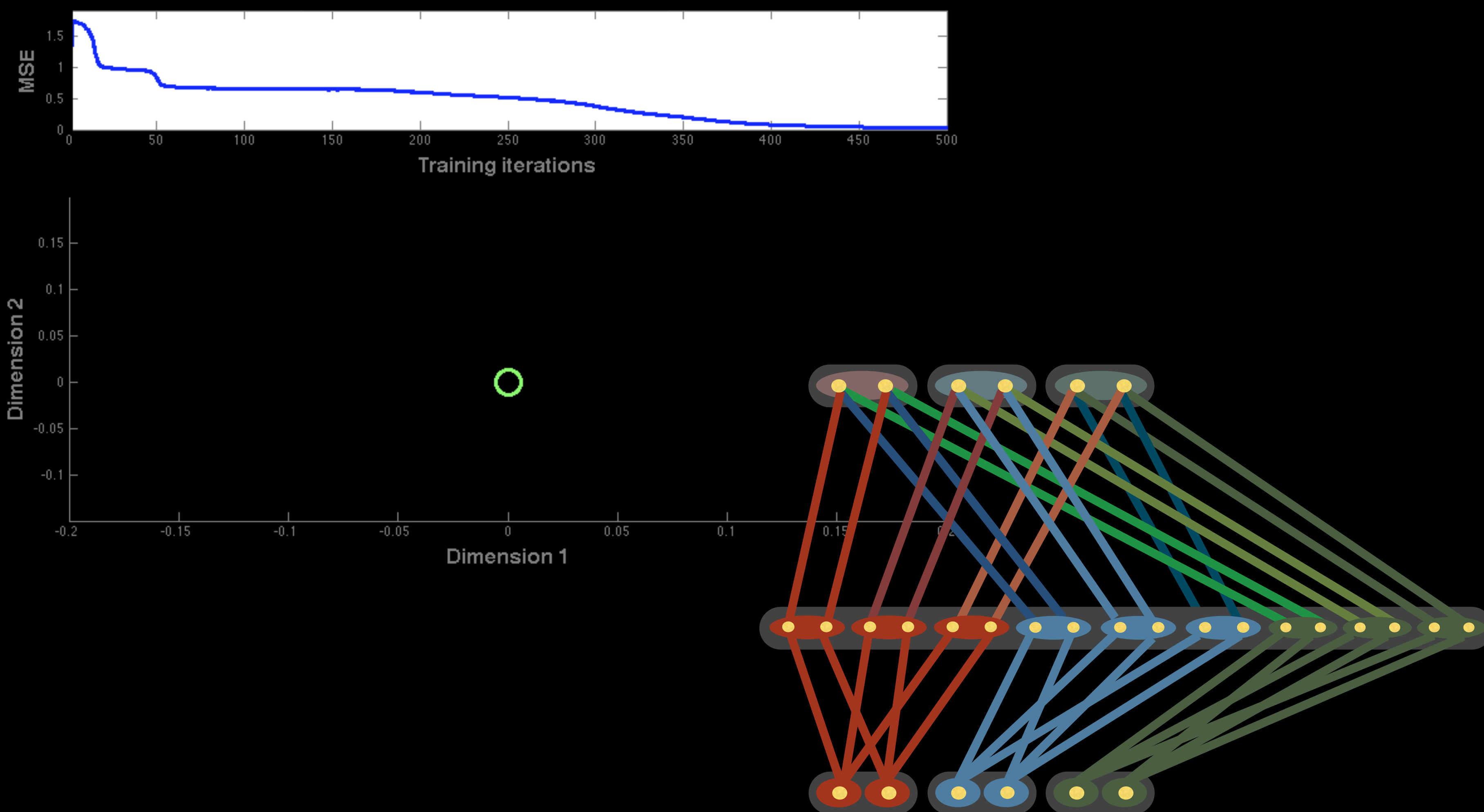




Single + Multitask Training



Single + Multitask Training



Example

JANUARY, 1923 SCIENTIFIC AMERICAN 17

Doing Two Things at Once

Multiple Consciousness, or Reflex Action of Unaccustomed Range?

By Dr. Alfred Gradenwitz

Thea Alba

Example

Doing Two Things at Once

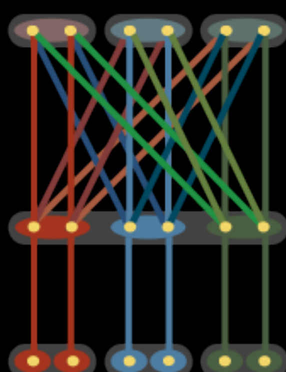
Multiple Consciousness, or Reflex Action of Unaccustomed Range?

By Dr. Alfred Gradenwitz

Thea Alba

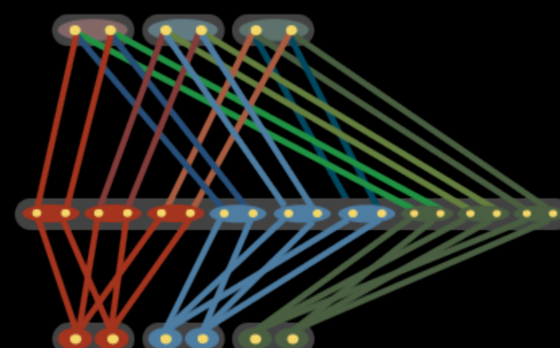
Empirical Implications

Multipurpose but serial



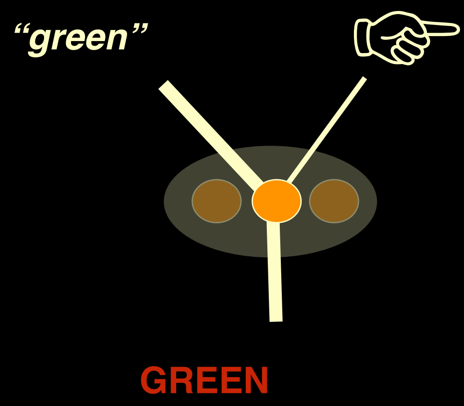
Learning

Multitasking and parallel



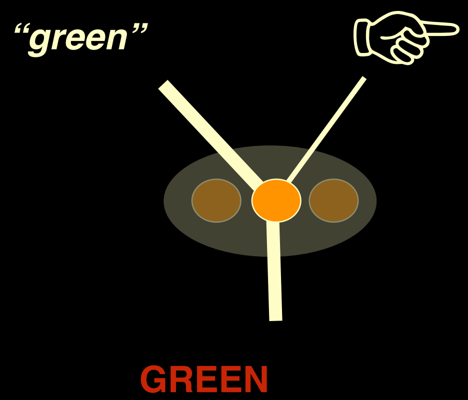
Empirical Study

Shared Representation
(can't* multitask with color naming*)

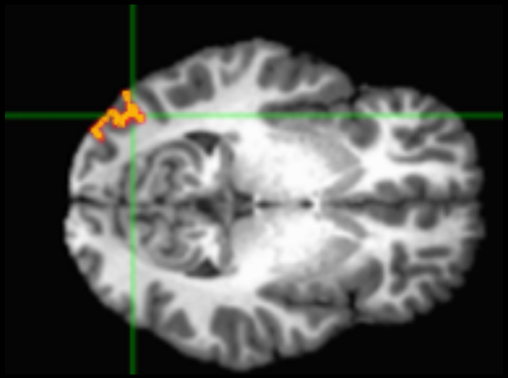


Empirical Study

Shared Representation
***(can't** multitask with color naming)*

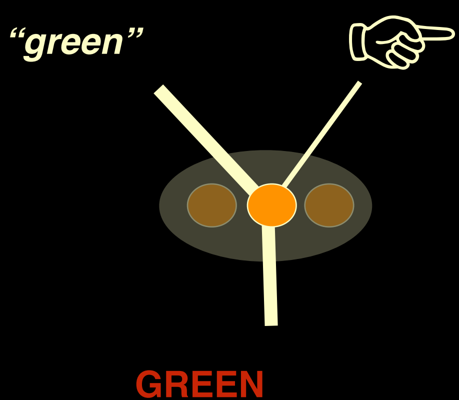


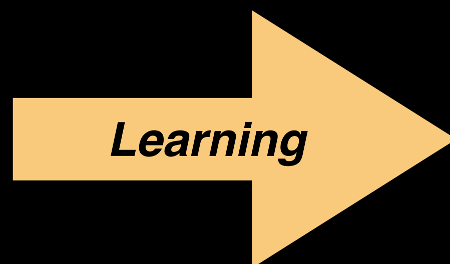
Predicted results



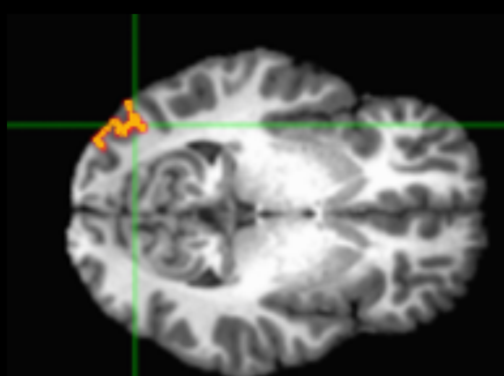
Empirical Study

Shared Representation
***(can’t** multitask with color naming)*





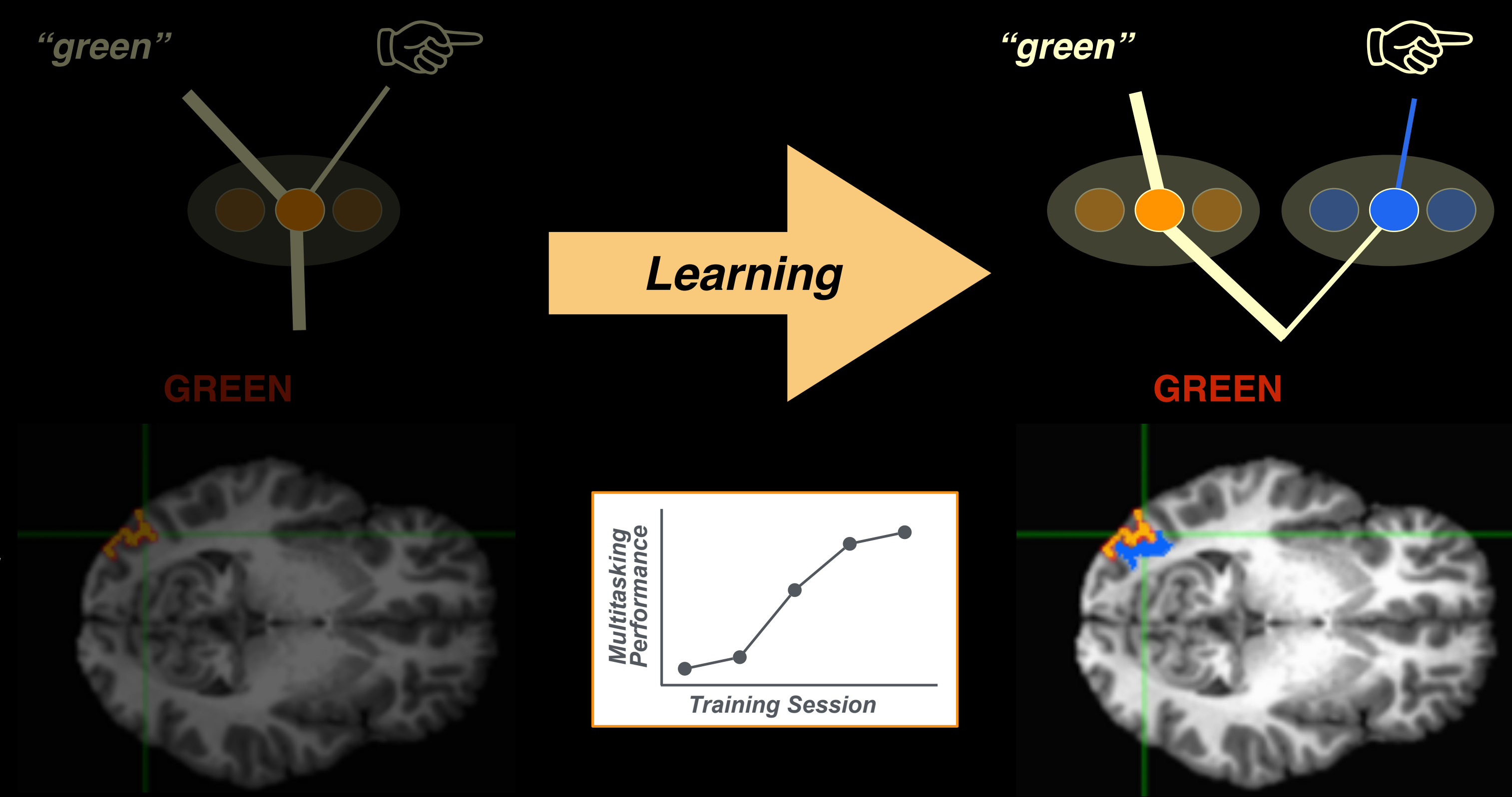
Predicted results



Empirical Study

Shared Representation
(*can't* multitask with color naming)

Separated Representation
(*can* multitask with color naming)



Predicted results

Summary

Summary

- Modest amounts of cross-talk produce a radical constraint on parallelism:
  - Need to go “serial” with even modest pathway overlap

    ⇒constraints on multitasking capacity

Summary

- Modest amounts of cross-talk produce a radical constraint on parallelism:
  - Need to go "serial" with even modest pathway overlap

    ⇒constraints on multitasking capacity

- Control architecture is adapted to this constraint:
  - reflects optimization of tradeoff between *shared* of representations and *processing efficiency*

Learning vs. Performance

Learning vs. Performance

- **Fundamental tradeoff:**

SHARED representations
generalization & flexibility:
learning efficiency

vs.

SEPARATED representations
multitasking:
performance efficiency

Learning vs. Performance

- **Fundamental tradeoff:**

BIAS

SHARED **representations**

generalization & flexibility: ***vs.***

learning **efficiency**

Learning vs. Performance

- **Fundamental tradeoff:**

  BIAS

  SHARED **representations**

  generalization & flexibility: ***vs.***

  learning **efficiency**

- **Explains:**
  - **Capacity constraints in cognitive control:**
    - *purpose of cognitive control rather than a limitation*

Learning vs. Performance

- **Fundamental tradeoff:**

SHARED **representations**
generalization & flexibility:
learning **efficiency**

vs.

- **Explains:**
  - **Capacity constraints in cognitive control:**
    - *purpose of cognitive control rather than a limitation*
  - **Continuum of serial vs. parallel processing in distributed systems**

Learning vs. Performance

- **Fundamental tradeoff:**

AUTOMATIZATION

SEPARATED **representations**

multitasking:

performance **efficiency**

- **Explains:**
  - **Capacity constraints in cognitive control:**
    - *purpose of cognitive control rather than a limitation*
  - **Continuum of serial vs. parallel processing in distributed systems**
  - **Trajectory from controlled to automatic processing (e.g., in skill acquisition)**

Learning vs. Performance

- Fundamental tradeoff:

AUTOMATIZATION

SEPARATED representations
multitasking:
performance efficiency

- Explains:
  - Capacity constraints in cognitive control:
    - *purpose of cognitive control rather than a limitation*
  - Continuum of serial vs. parallel processing in distributed systems
  - Trajectory from controlled to automatic processing (e.g., in skill acquisition)
  - Other fundamental psychological phenomena:
    - *attention & "binding"*

Learning vs. Performance

- **Fundamental tradeoff:**

AUTOMATIZATION

SEPARATED **representations**

multitasking:

performance **efficiency**

- **Explains:**
  - **Capacity constraints in cognitive control:**
    - *purpose of cognitive control rather than a limitation*
  - **Continuum of serial vs. parallel processing in distributed systems**
  - **Trajectory from controlled to automatic processing (e.g., in skill acquisition)**
  - **Other fundamental psychological phenomena:**
    - *attention & "binding"*
  - **Defines an intertemporal choice between "getting by" and "getting trained"**
    - *how does the system decide this?*
    - *how can this be formalized?*

Fundamental Tension

- *Interactive* Parallelism

- *Independent* Parallelism

Fundamental Tension

- *Interactive* Parallelism
  - many small *interacting* computations in the service of some *single* coherent higher level process

- *Independent* Parallelism

Fundamental Tension

- *Interactive* Parallelism
  - at the core of *PDP* (and now deep learning)
- *Independent* Parallelism

Fundamental Tension

- *Interactive* Parallelism

  – relies on *shared representations* *(learning efficiency)*

- *Independent* Parallelism

Fundamental Tension

- *Interactive* Parallelism
  - requires "protection" from interference ⇒ *control-dependent*
- *Independent* Parallelism

Fundamental Tension

- *Interactive* Parallelism
  - requires “protection” from interference ⇒ *control-dependent*
- *Independent* Parallelism
  - many *separate* unrelated computations

Fundamental Tension

- *Interactive* Parallelism
  - requires "protection" from interference ⇒ *control-dependent*
- *Independent* Parallelism
  - at the core of traditional *"embarrassing" parallelism* (e.g., MPI)

Fundamental Tension

- *Interactive* Parallelism
  - requires "protection" from interference ⇒ *control-dependent*
- *Independent* Parallelism
  - relies on *separated representations*

Fundamental Tension

- *Interactive* Parallelism
  - requires “protection” from interference ⇒ *control-dependent*
- *Independent* Parallelism
  - allows “multitasking” ⇒ *automaticity* *(processing efficiency)*

Outline

- Multitasking and Control
- ☞ Miller's Law

VOL. 63, NO. 2 MARCH, 1956

THE PSYCHOLOGICAL REVIEW

THE MAGICAL NUMBER SEVEN, PLUS OR MINUS TWO: SOME LIMITS ON OUR CAPACITY FOR PROCESSING INFORMATION[1]

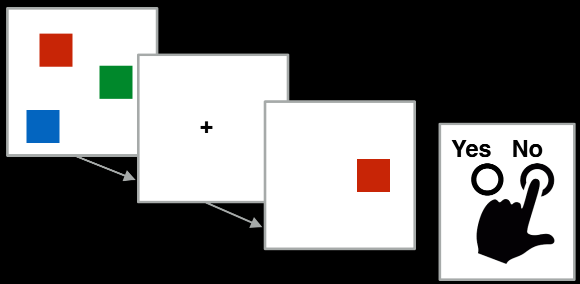
GEORGE A. MILLER

Harvard University

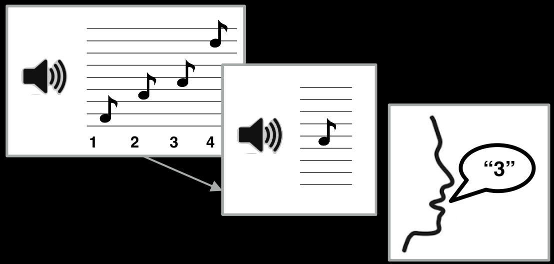
My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable. The persistence with which this number plagues me is far more than a random accident. There is, to quote a famous senator, a design behind it, some pattern governing its appearances. Either there really is something unusual about the number or else I am suffering from delusions of persecution. . . .

3 Tasks

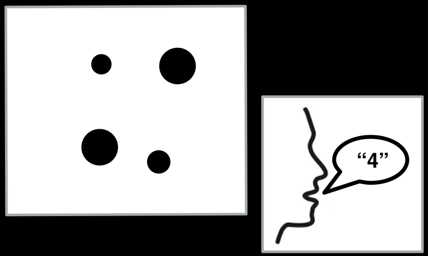
- Visual Working Memory Task



- Absolute Perceptual Judgement

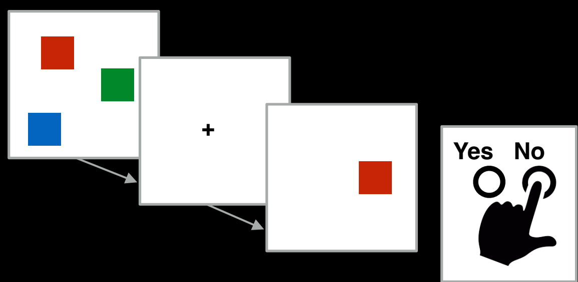


- Numerosity Estimation ("subtilizing")

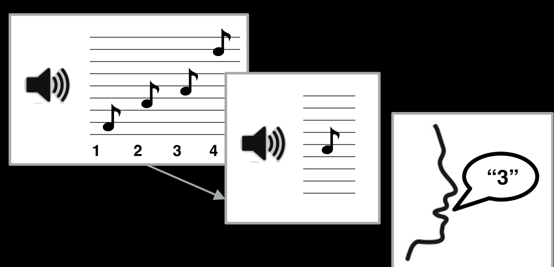


3 Tasks

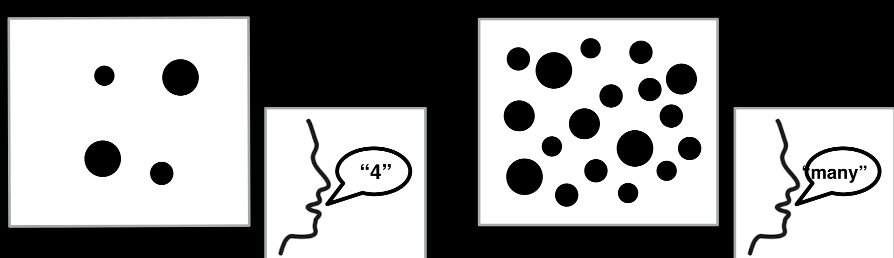
- Visual Working Memory Task



- Absolute Perceptual Judgement

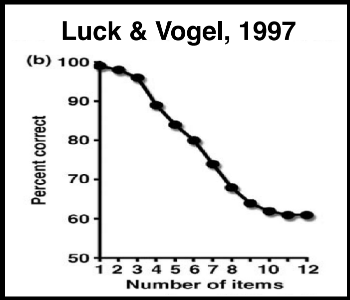


- Numerosity Estimation (“subtilizing”)

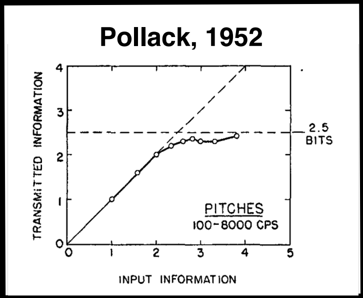


3 Tasks

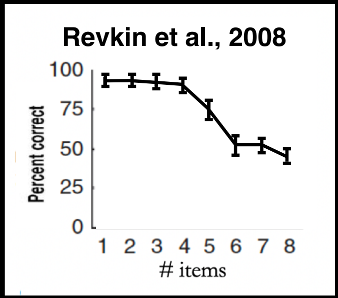
- Visual Working Memory Task



- Absolute Perceptual Judgement

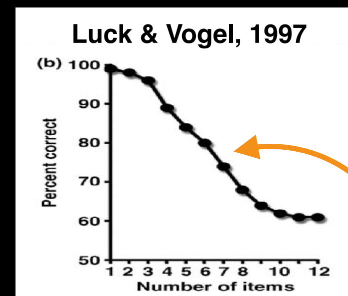


- Numerosity Estimation (“subtilizing”)

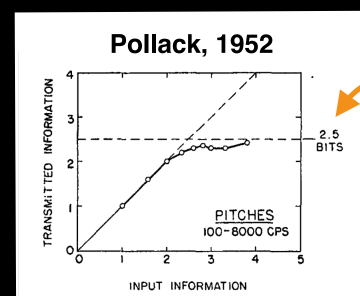


3 Tasks

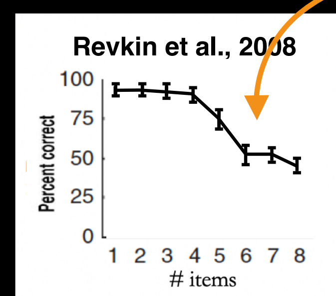
- Visual Working Memory Task



- Absolute Perceptual Judgement



- Numerosity Estimation ("subtilizing")



7 +/- 2

(~2.5 bits)

Vol. 63, No. 2 March, 1956

THE PSYCHOLOGICAL REVIEW

THE MAGICAL NUMBER SEVEN, PLUS OR MINUS TWO: SOME LIMITS ON OUR CAPACITY FOR PROCESSING INFORMATION [1]

GEORGE A. MILLER

Harvard University

My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable. The persistence with which this number plagues me is far more than a random accident. There is, to quote a famous senator, a design behind it, some pattern governing its appearances. Either there really is something unusual about the number or else I am suffering from delusions of persecution. . . .

. . . For the present I propose to withhold judgment. Perhaps there is something deep and profound behind all these sevens, something just calling out for us to discover it. But I suspect that it is only a pernicious, Pythagorean coincidence.

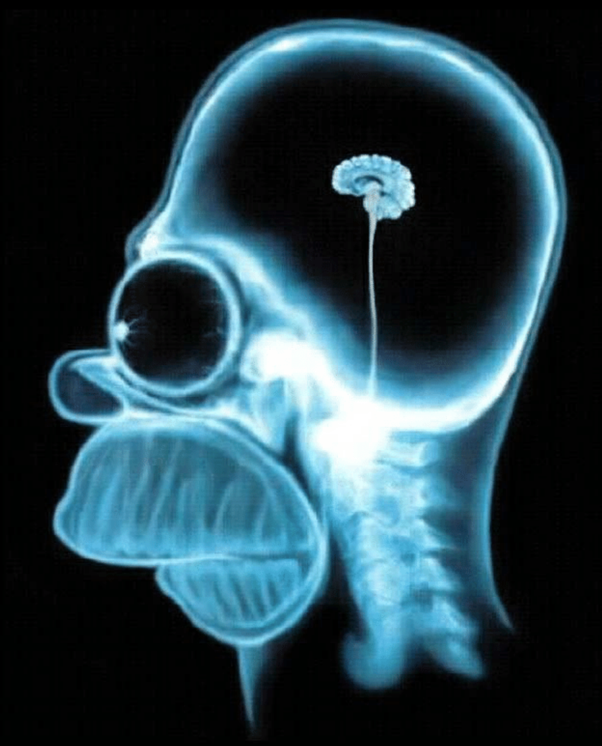
And the list goes on…

- **Classic verbal short-term memory task**
  (Sternberg, 1966)

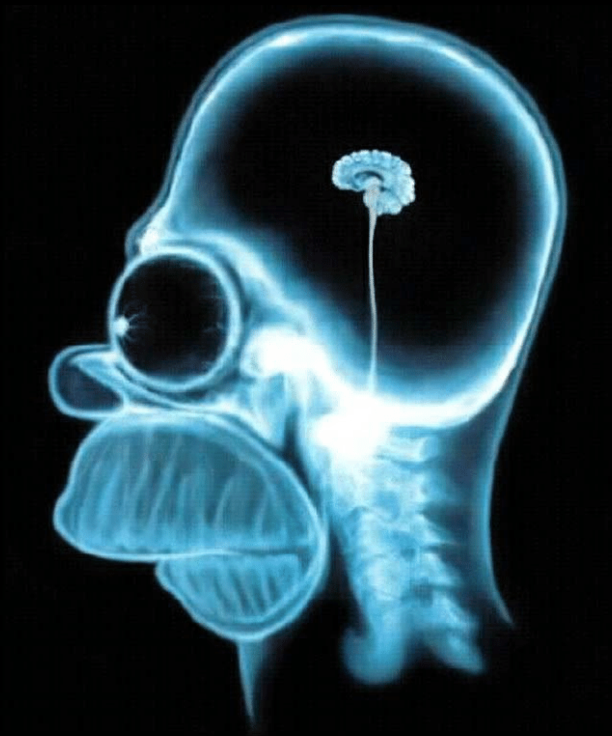
- **Classic working memory tasks**
  (Baddeley, 1990; Cowan, 1999; Luck & Vogel, 1997)

- **Attention / visual search tasks**
  (Shiffrin & Schneider, 1977; Treisman & Gelade, 1980)

- **Control-dependent processing**
  (Posner & Snyder,1975; Pashler, 1994)



And the list goes on…



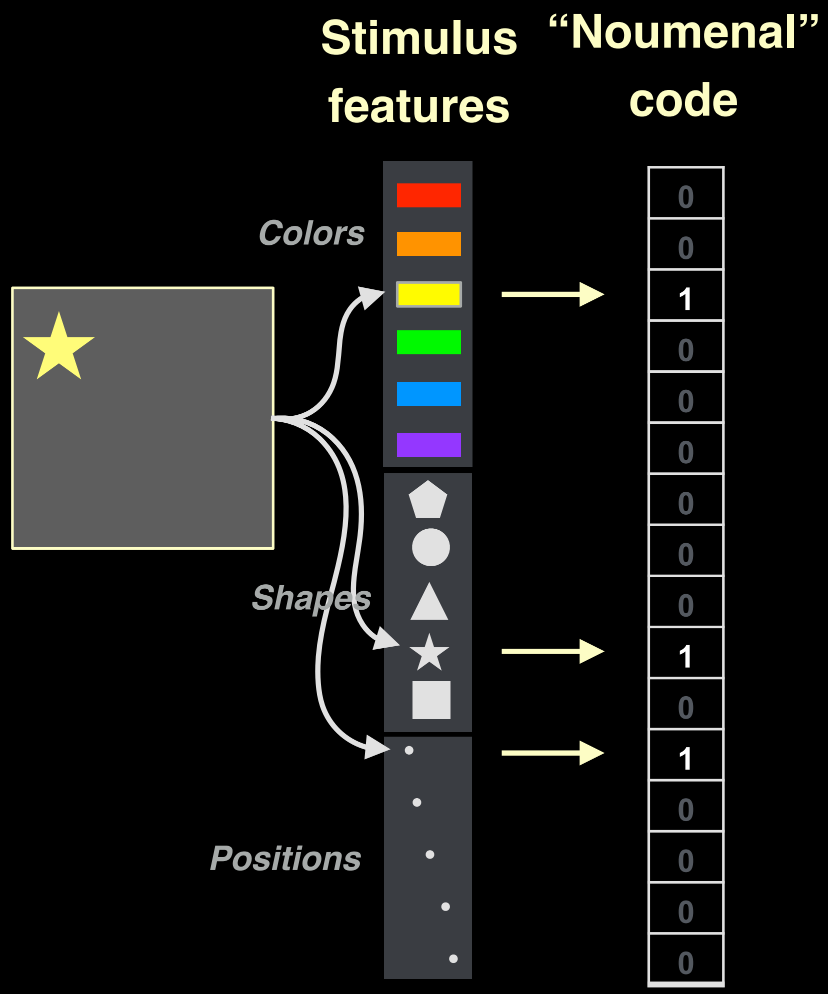
- All exhibit ≤ 2½ bit capacity limit

And the list goes on…

- But for the same reason?

An Information Theoretic Perspective

Empirical State



An Information Theoretic Perspective

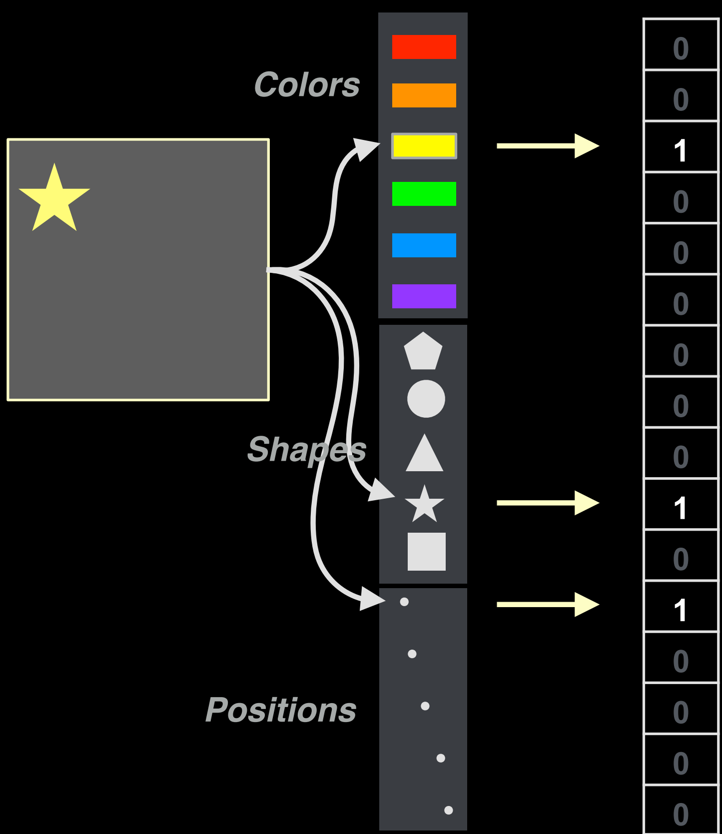
Empirical State

Internal Representation

Stimulus features

"Noumenal" code

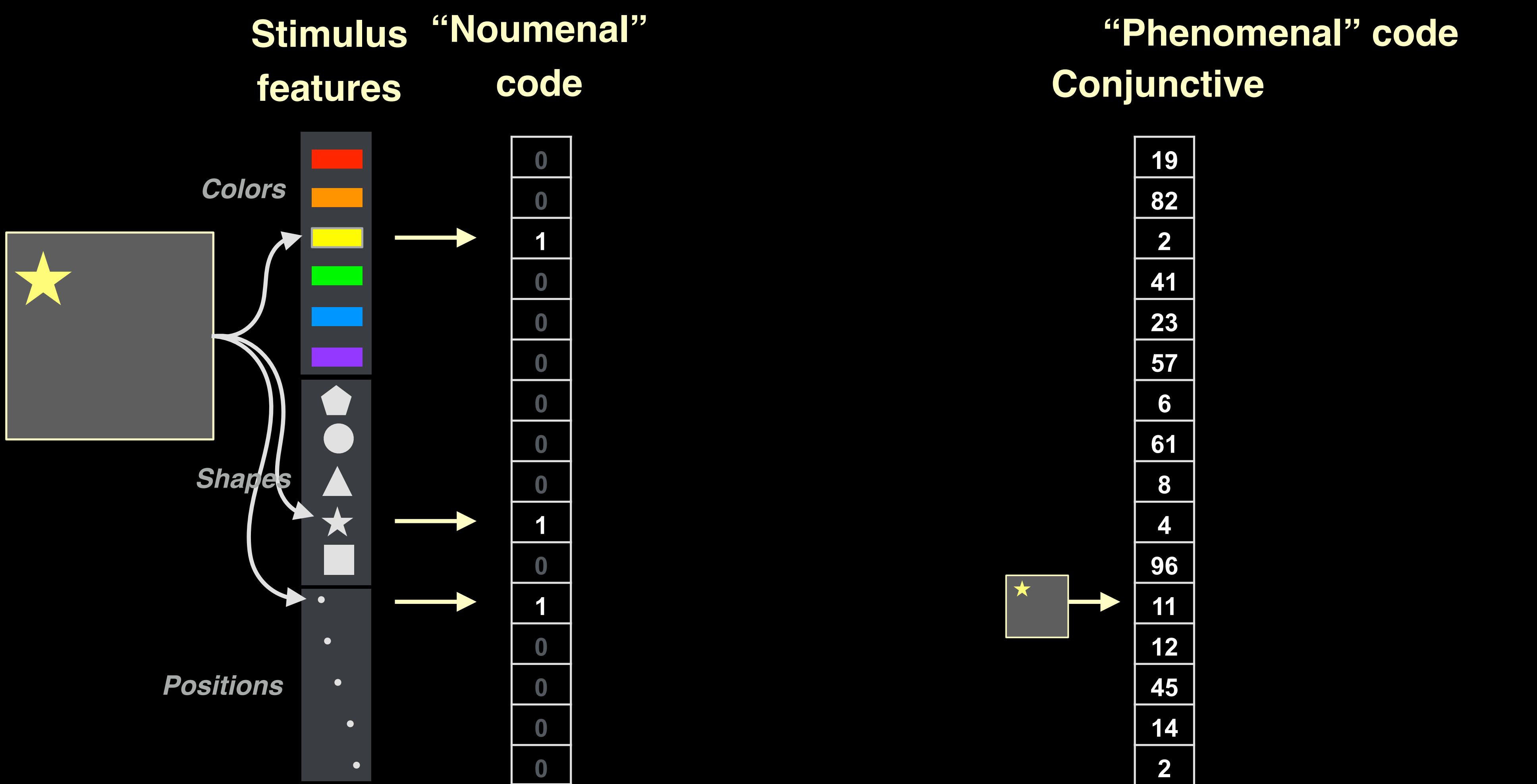
"Phenomenal" code



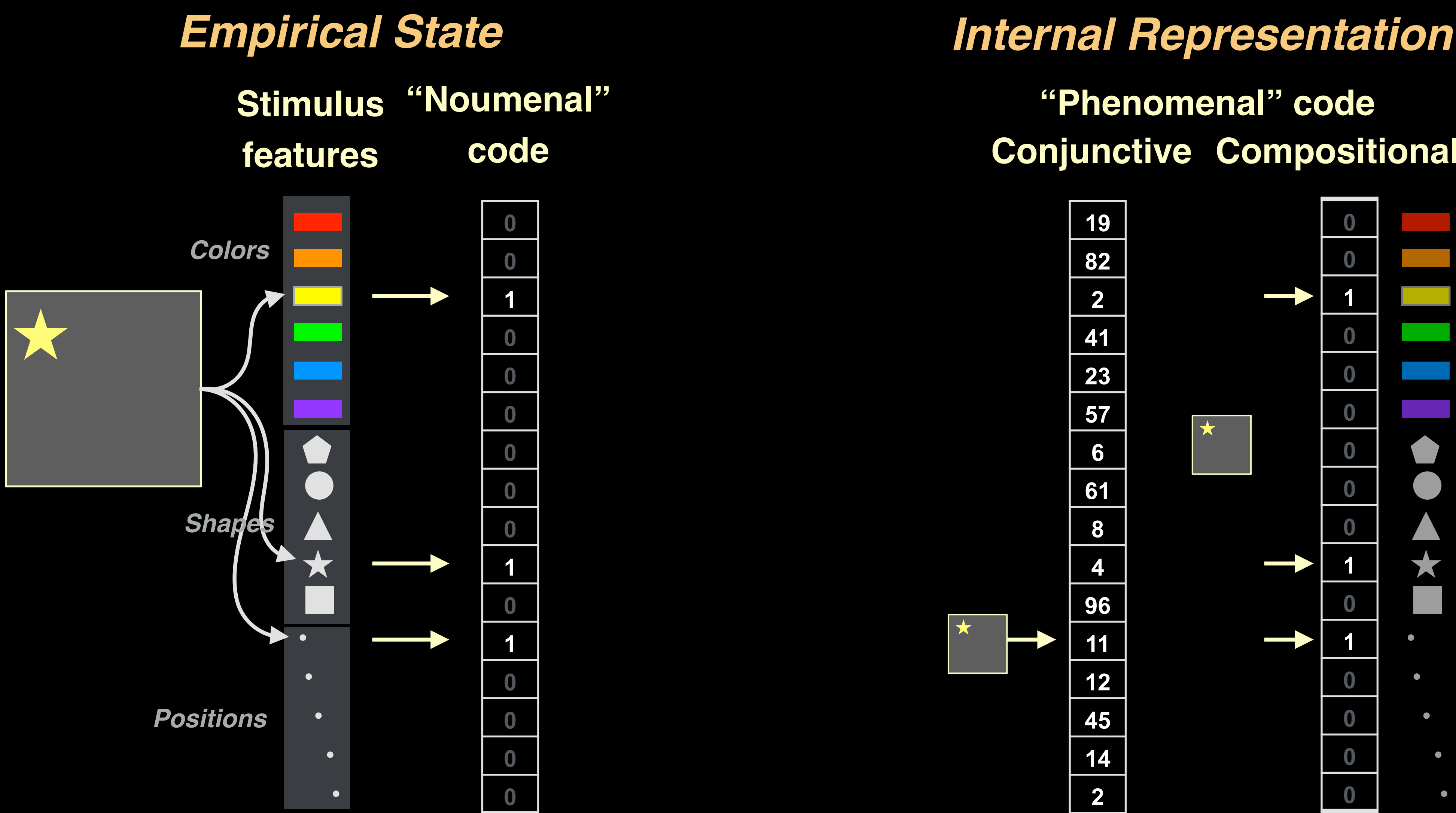
An Information Theoretic Perspective

Empirical State

Internal Representation



An Information Theoretic Perspective



An Information Theoretic Perspective

Empirical State

Internal Representation

"Noumenal" code	"Phenomenal" code: Conjunctive	"Phenomenal" code: Compositional
0	19	0
0	82	0
1	2	1
0	41	0
0	23	0
0	57	0
0	6	0
0	61	0
0	8	0
1	4	1
0	96	0
1	11	1
0	12	0
0	45	0
0	14	0
0	2	0

—— *Transmission* ⟶

***Capacity*:**

⟵ **Mutual Information** ⟶

$H(X) + H(Y) - H(X,Y)$

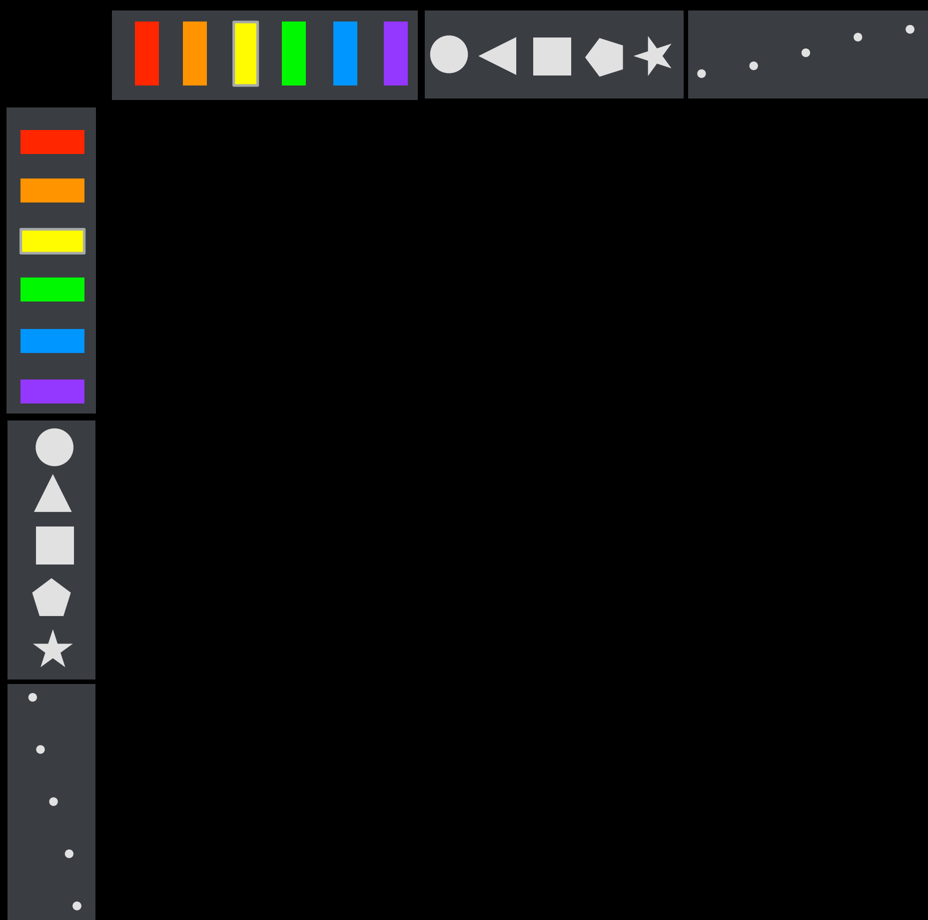
Semanticity of Codes

Conjunctive Code

1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Semanticity of Codes

Conjunctive Code



Semantic + Compositional Code

1	.3	.1	0	0	.1	0	0	0	0	0	0	0	0	0	0
.3	1	.3	.1	0	0	0	0	0	0	0	0	0	0	0	0
.1	.3	1	.3	.1	0	0	0	0	0	0	0	0	0	0	0
0	.1	.3	1	.3	.1	0	0	0	0	0	0	0	0	0	0
0	0	.1	.3	1	.3	0	0	0	0	0	0	0	0	0	0
.1	0	0	.1	.3	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	.1	0	.2	0	0	0	0	0	0
0	0	0	0	0	0	.1	1	.8	.4	.3	0	0	0	0	0
0	0	0	0	0	0	0	.8	1	.1	.1	0	0	0	0	0
0	0	0	0	0	0	.2	.4	.1	1	.9	0	0	0	0	0
0	0	0	0	0	0	0	.3	.1	.9	1	0	0	0	0	0
0	0	0	0	0	0	0	0	.4	.6	0	1	.8	.6	.4	.2
0	0	0	0	0	0	0	0	0	.4	0	.8	1	.8	.6	.4
0	0	0	0	0	0	0	0	0	0	0	.6	.8	1	.8	.6
0	0	0	0	0	0	0	0	0	0	0	.4	.6	.8	1	.8
0	0	0	0	0	0	0	0	0	0	0	.2	.4	.6	.8	1

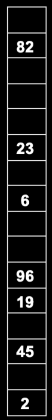
Two Types of Error

Two Types of Error

- **Identification error:**
  - **Traditional** ***information theoretic*** **construct**
    (e.g, Optimal Coding, Rate Distortion Theory, etc.)
  - **Loss:** transmitted code is ***not*** *same as* source code

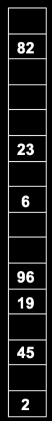
Two Types of Error

- **Identification error:**
  - **Traditional** ***<u>information theoretic</u>*** **construct**
    (e.g, Optimal Coding, Rate Distortion Theory, etc.)
  - **Loss:** transmitted code is ***not*** *<u>same as</u>* source code
  - Minimized by maximizing distance between codes:
    ⇒ conjunctive codes



Two Types of Error

- **Identification error:**
  - **Traditional** *information theoretic* **construct** *(e.g, Optimal Coding, Rate Distortion Theory, etc.)*
  - **Loss:** transmitted code is ***not*** *same as* source code
  - Minimized by maximizing distance between codes: ⇒ **conjunctive codes**
  - Probed by **accuracy** tasks



Two Types of Error

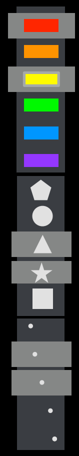
- **Identification error:**
  - **Traditional *information theoretic* construct**
    (e.g, Optimal Coding, Rate Distortion Theory, etc.)
  - **Loss:** transmitted code is ***not*** *same as* source code
  - Minimized by maximizing distance between codes:
    ⇒ **conjunctive codes**
  - Probed by **accuracy** tasks

- **Generalization error:**
  - Traditional *cognitive* / *modern ML* construct
    (e.g, Shepherd's Universal Principle of Generalization)
  - **Loss**: transmitted code is **not** *close to* source code

Two Types of Error

- **Identification error:**
  - **Traditional** ***information theoretic*** **construct**
    (e.g, Optimal Coding, Rate Distortion Theory, etc.)
  - **Loss:** transmitted code is ***not*** *same as* source code
  - Minimized by maximizing distance between codes:
    ⇒ **conjunctive codes**
  - Probed by **accuracy** tasks

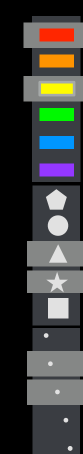
- **Generalization error:**
  - Traditional *cognitive* / *modern ML* construct
    (e.g, Shepherd's Universal Principle of Generalization)
  - **Loss**: transmitted code is **not** *close to* source code
  - Minimized by matching codes to structure of data:
    ⇒ **semantic / compositional codes**



Two Types of Error

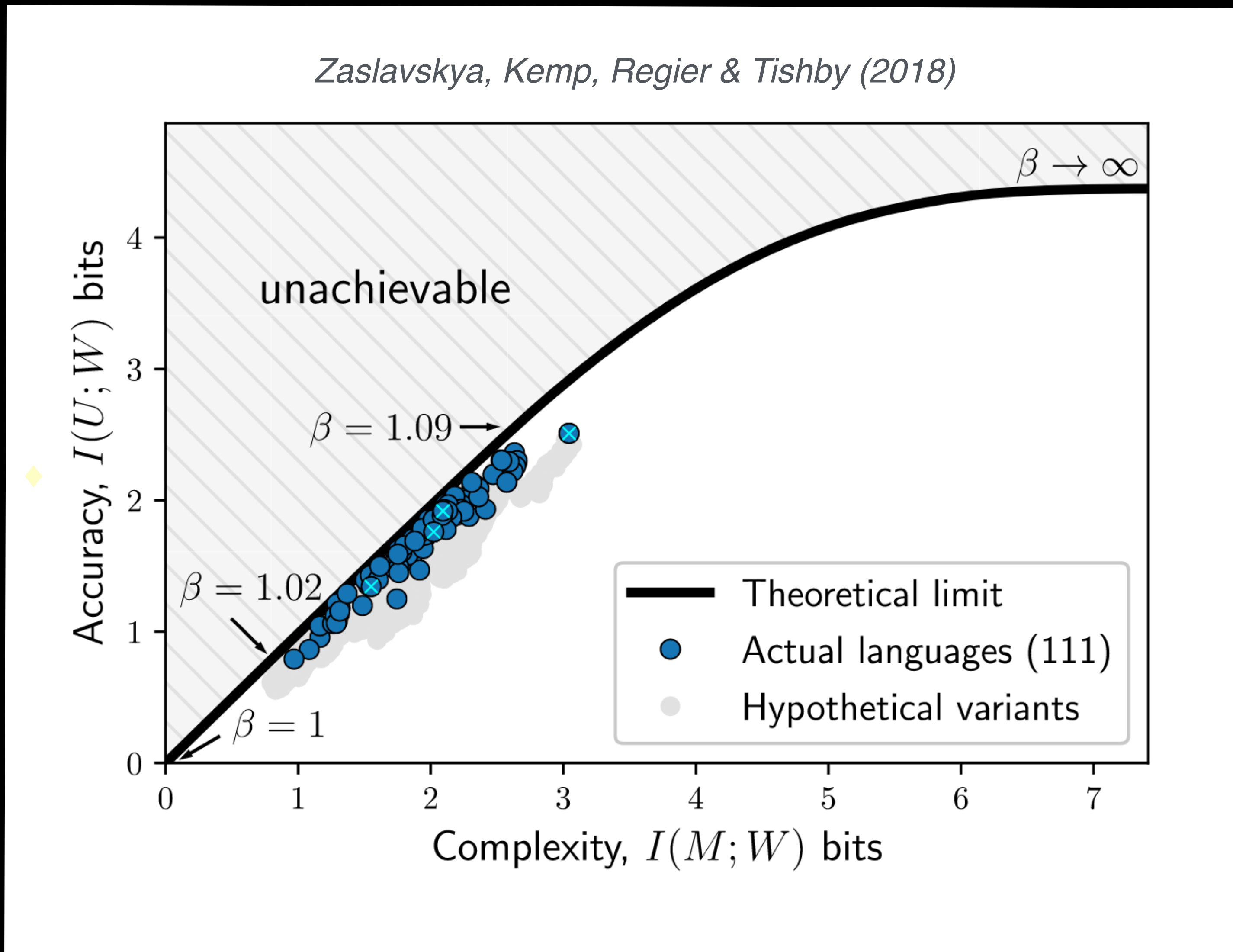
- **Identification error:**
  - **Traditional** ***information theoretic*** **construct**
    (e.g, Optimal Coding, Rate Distortion Theory, etc.)
  - **Loss:** transmitted code is ***not same as*** source code
  - Minimized by maximizing distance between codes:
    ⇒ **conjunctive codes**
  - Probed by **accuracy** tasks

- **Generalization error:**
  - Traditional *cognitive* / *modern ML* construct
    (e.g, Shepherd's Universal Principle of Generalization)
  - **Loss**: transmitted code is **not** *close to* source code
  - Minimized by matching codes to structure of data:
    ⇒ **semantic / compositional codes**
  - Probed by **similarity** tasks



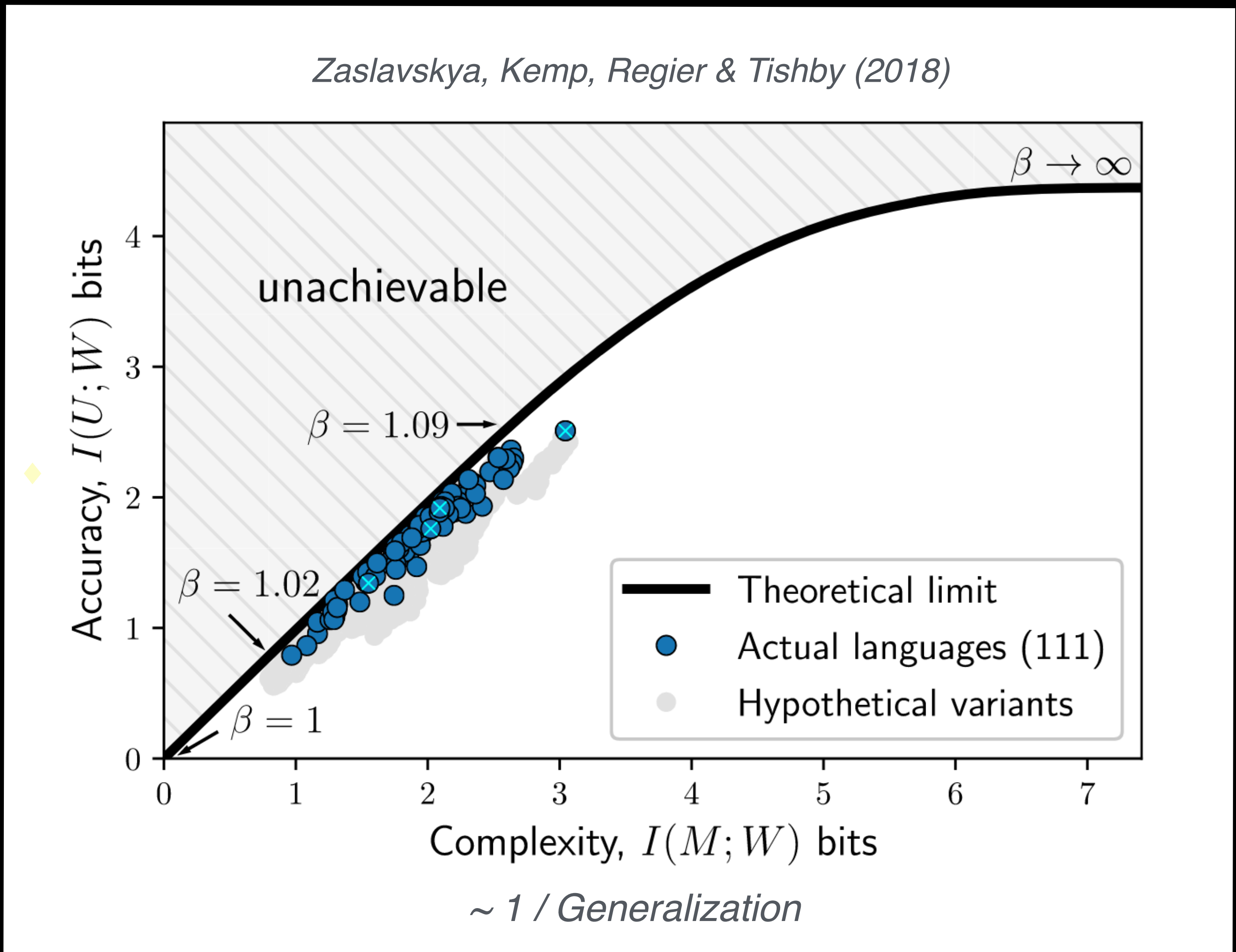
Information Bottleneck

Tishby, Naftali, Pereira & William Bialek (2000)



Information Bottleneck

Tishby, Naftali, Pereira & William Bialek (2000)



Abstract (Numerical) Model

Abstract (Numerical) Model

- **Assume some covariance matrix over features of codes**
  - implement similarity structure as exponentially decaying distance between features within a dimension (Shepherd, 1987)

Abstract (Numerical) Model

- **Assume some covariance matrix over features of codes**
  - implement similarity structure as exponentially decaying distance between features within a dimension (Shepherd, 1987)
- **Model two tasks:**
  - **Similarity task:**

    indicate stimulus in a set *most similar* to randomly selected probe

    indexes *representational capacity* *(as a function of coding structure)*

Abstract (Numerical) Model

- **Assume some covariance matrix over features of codes**
  - implement similarity structure as exponentially decaying distance between features within a dimension (Shepherd, 1987)
- **Model two tasks:**
  - **Similarity task:**

    indicate stimulus in a set *most similar* to randomly selected probe

    indexes *representational capacity* *(as a function of coding structure)*
  - **Identification task:**

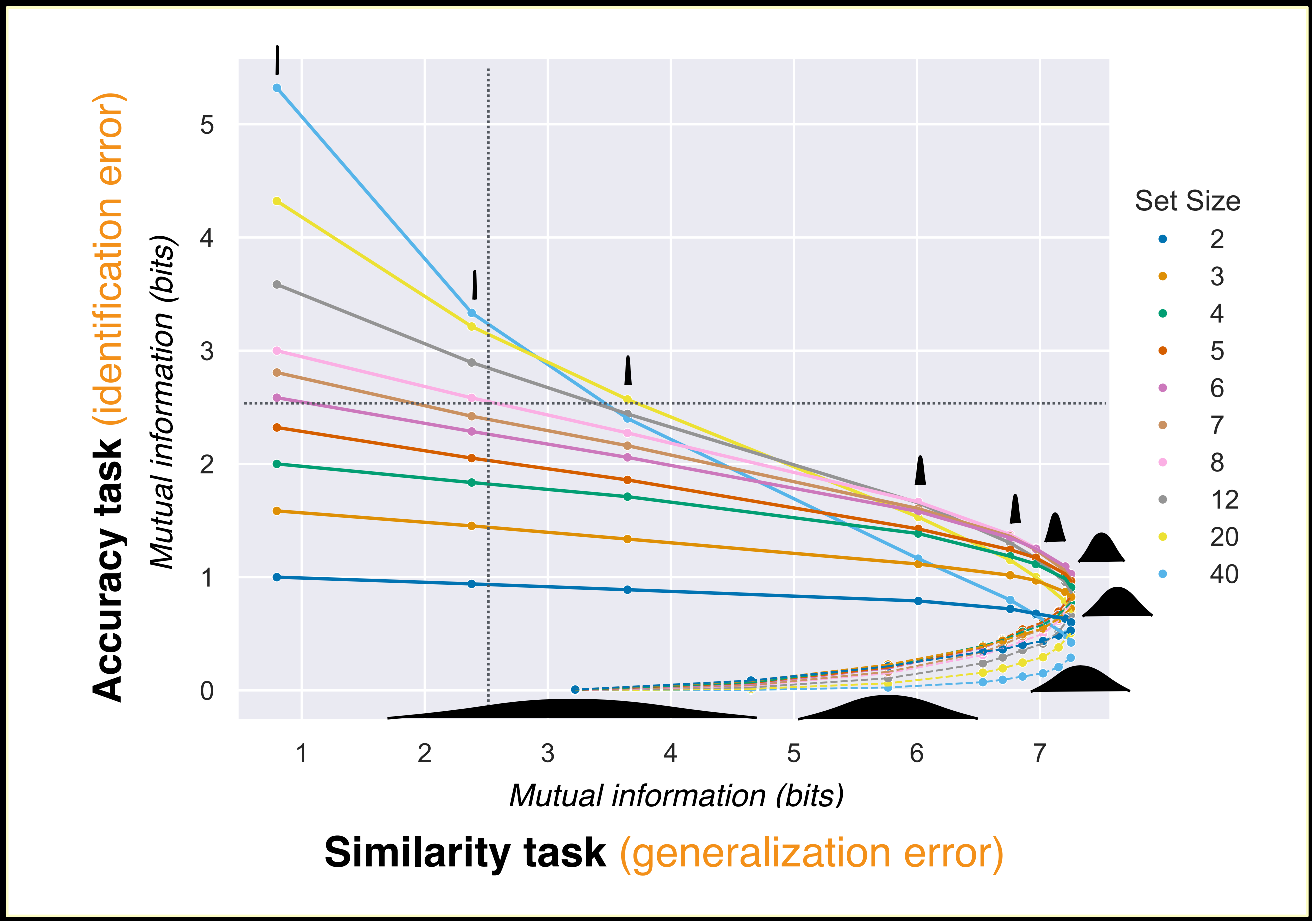
    identify probed item (based on one of its features)

    indexes *processing capacity* *(as a function of # items = set size)*

Miller's Law

Frankland, Webb, Lewis & Cohen (2025)

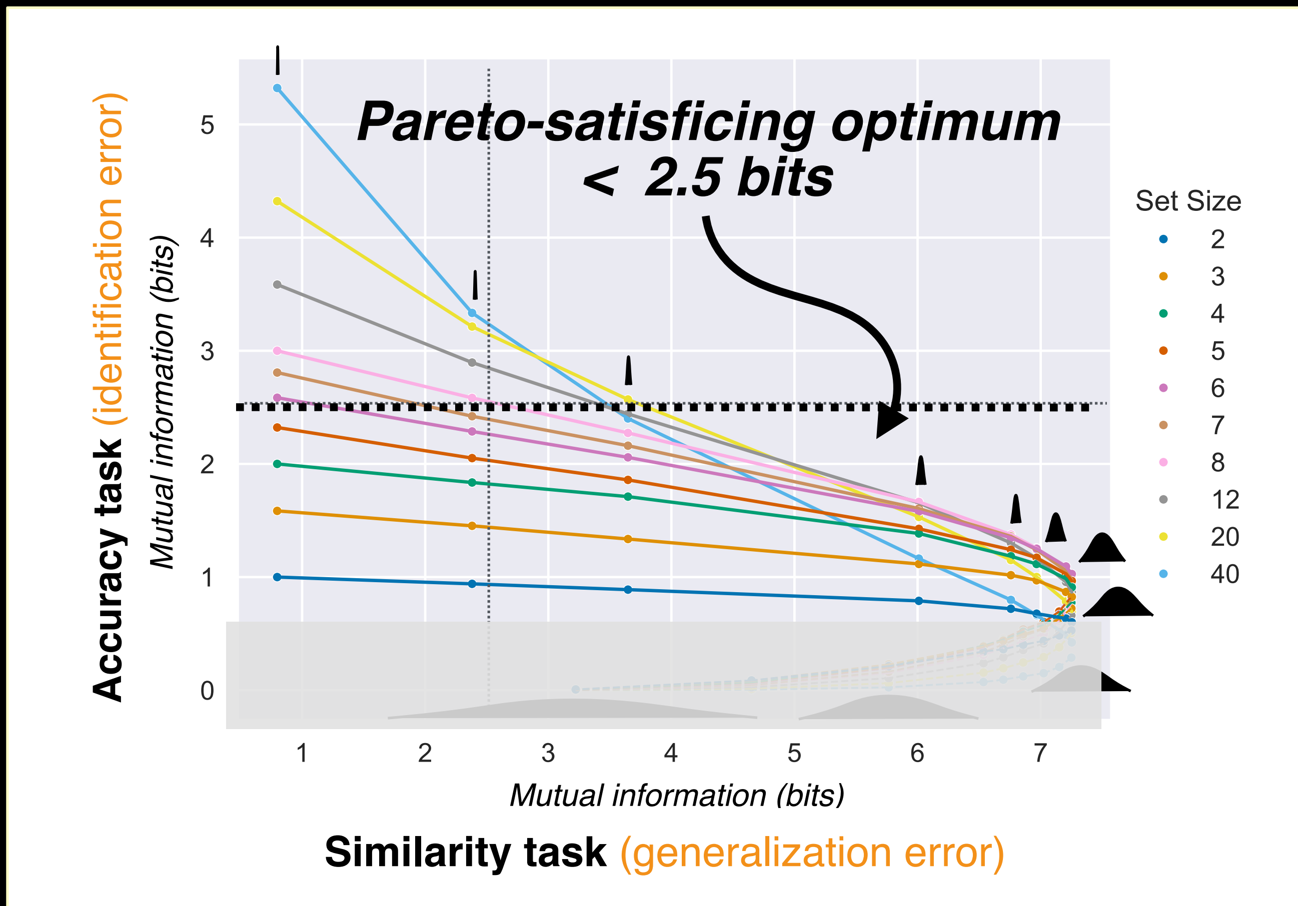
Identification Error ~1/∝ Generalization Error



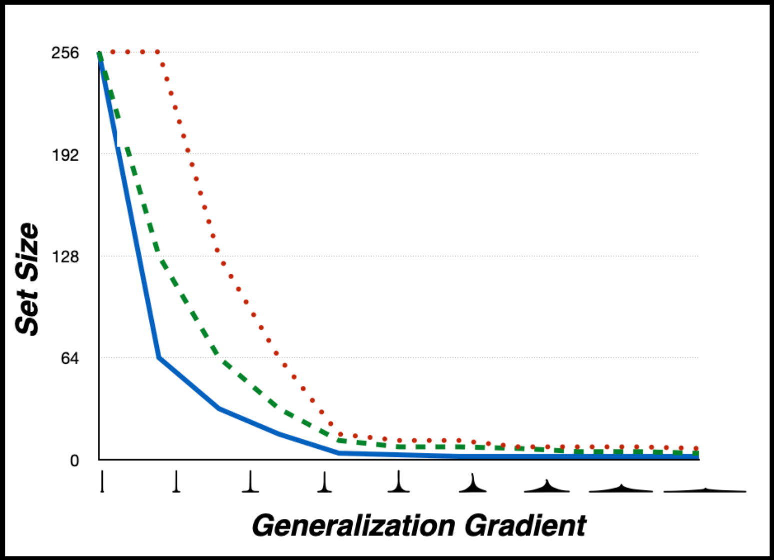
Miller’s Law

Frankland, Webb, Lewis & Cohen (2025)

Identification Error ~1/∝ Generalization Error



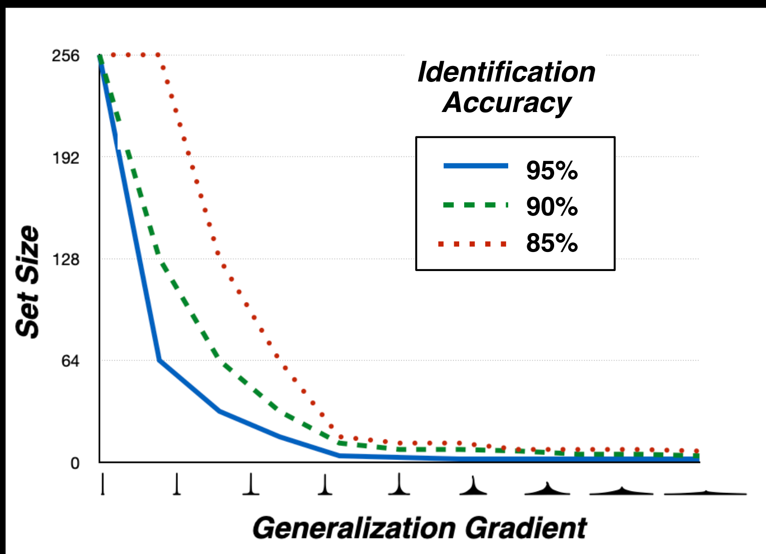
Ubiquitous Effect



Ubiquitous Effect

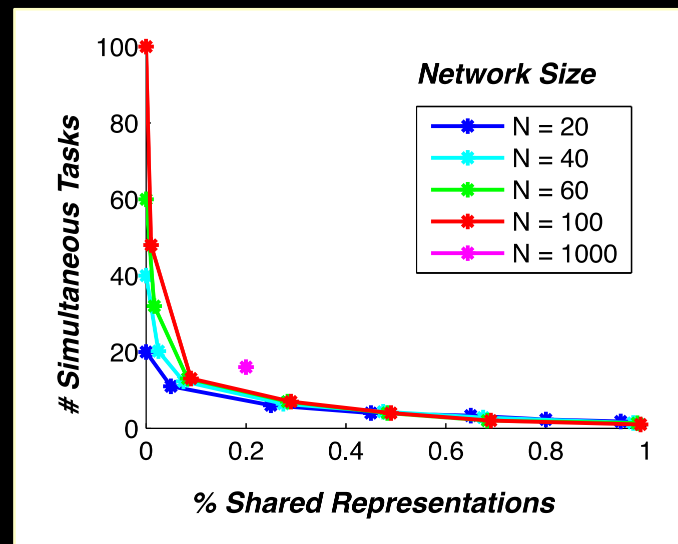
Inference

(Frankland et al., 2024)

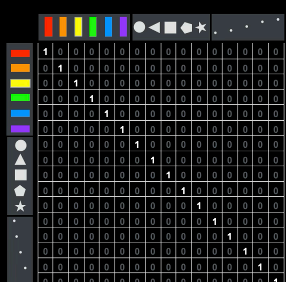


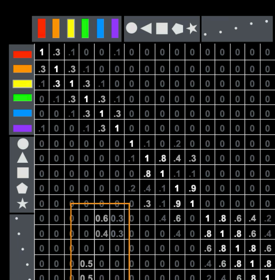
Affordance

(Musslick et al., 2023)

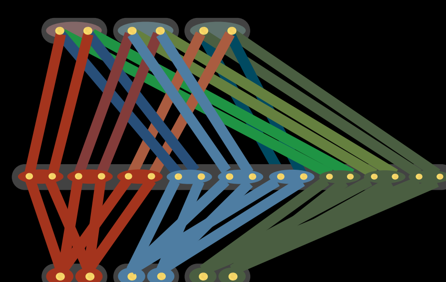


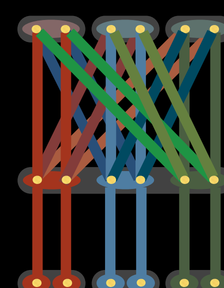
Conjunctive *vs.* ***Conjunctive***





Conjunctive *vs.* ***Conjunctive***





Fundamental Principle

Fundamental Principle

- **Inescapable tradeoff between:**
  - **accuracy (identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)

Fundamental Principle

- **Inescapable tradeoff between:**
  - **accuracy (identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)
  - **similarity and abstraction (generalization error)**
    - demands that similar things be represented similarly
      (Shepherd's Universal Law of Generalization)

Fundamental Principle

- **Inescapable tradeoff between:**
  - **accuracy (identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)
  - **similarity and abstraction (generalization error)**
    - demands that similar things be represented similarly
      (Shepherd's Universal Law of Generalization)

Miller's Law:

*[**Representational** efficiency]^k • [**Processing** efficiency] $\propto$ c*

Fundamental Principle

- **Inescapable tradeoff between:**
  - **accuracy (identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)
  - **similarity and abstraction (generalization error)**
    - demands that similar things be represented similarly
      (Shepherd's Universal Law of Generalization)

Miller's Law:

*[**Representational** efficiency]^k • [**Processing** efficiency] ∝ c*

$$Error_{gen} \cdot \frac{e^{-\frac{1}{2}Error_{id}} + \sigma}{e^{-\frac{2}{3}Error_{id}} + \sigma} = \frac{1}{2}$$

Fundamental Principle

- **Inescapable tradeoff between:**
  - **accuracy (identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)
  - **similarity and abstraction (generalization error)**
    - demands that similar things be represented similarly
      (Shepherd's Universal Law of Generalization)

Miller's Law:

*[**Representational** efficiency]^k • [**Processing** efficiency] $\propto$ c*

$$Error_{gen} \cdot \frac{e^{-\frac{1}{2}Error_{id}} + \sigma}{e^{-\frac{2}{3}Error_{id}} + \sigma} = \frac{1}{2}$$

- **Optimum for generalization dramatically constrains identification:**

Fundamental Principle

- Inescapable tradeoff between:
  - **accuracy** **(identification error)**
    - demands as much precision as possible
      (Shannon's Rate-Distortion Theory)
  - **similarity and abstraction** **(generalization error)**
    - demands that similar things be represented similarly
      (Shepherd's Universal Law of Generalization)

Miller's Law:

*[**Representational** efficiency]^k • [**Processing** efficiency] $\propto$ c*

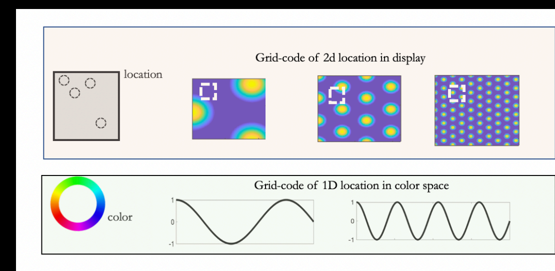
$$Error_{gen} \cdot \frac{e^{-\frac{1}{2}Error_{id}} + \sigma}{e^{-\frac{2}{3}Error_{id}} + \sigma} = \frac{1}{2}$$

- **Optimum for generalization**
  dramatically constrains identification:

- **Case study:** ***"Miller's tasks"***

Mechanistic Process Model

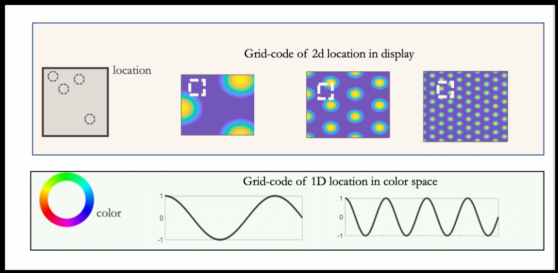
Steven Frankland
Dartmouth University



Mechanistic Process Model

Steven Frankland
Dartmouth University

- **Encoding**
  - *semantic, compositional* codes *(grid cells)*
  - informed by available *neural / psychophysical data:*

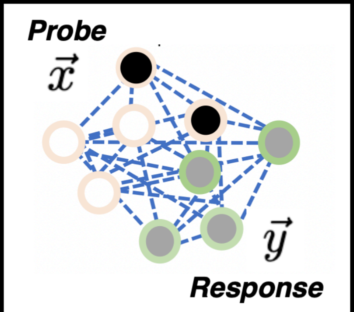


Mechanistic Process Model

Steven Frankland
Dartmouth University

- **Encoding**
  - *semantic, compositional* codes *(grid cells)*
  - informed by available *neural / psychophysical data:*

- **Memory**
  - rapid *Hebbian associative learning* in Hopfield network



Mechanistic Process Model

Steven Frankland
Dartmouth University

- **Encoding**
  - *semantic, compositional* codes *(grid cells)*
  - informed by available *neural / psychophysical data:*

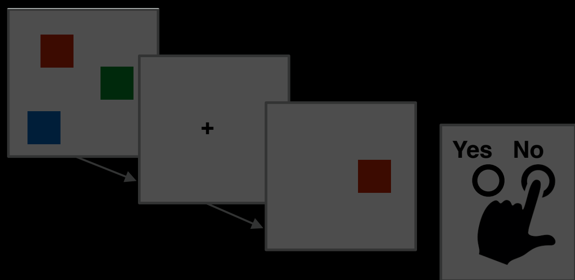
- **Memory**
  - rapid *Hebbian associative learning* in Hopfield network

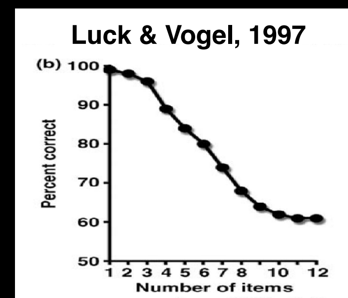
- **Processing**
  - embed items
  - encode in associative memory
  - clamp probe
  - settle to convergence (accuracy ≈ cross-entropy; cycles ≈ reaction time)
  - evaluate as a function of set size

Miller's Tasks

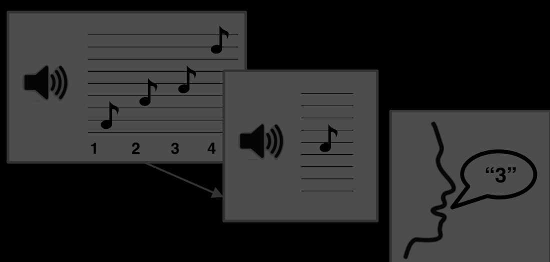
Data

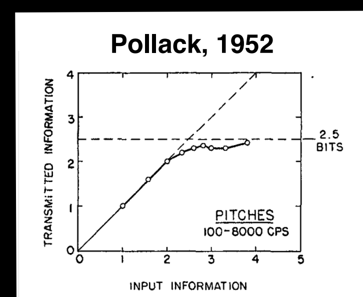
- Visual Working Memory Task



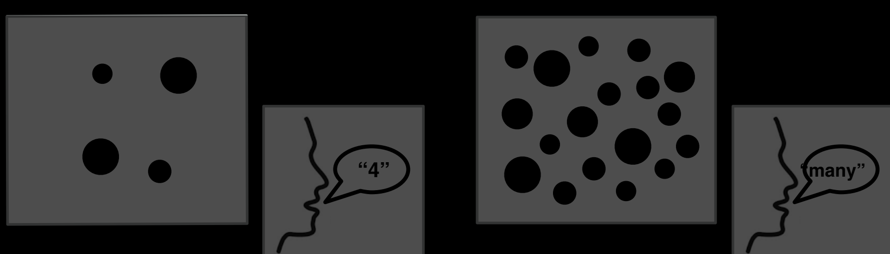


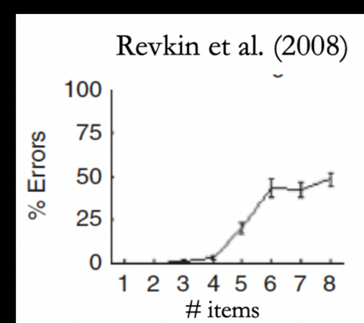
- Absolute Perceptual Judgement





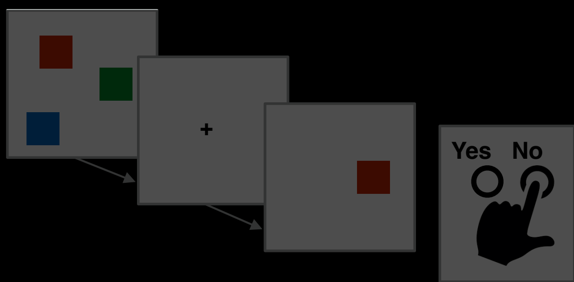
- Numerosity Estimation ("subtilizing")



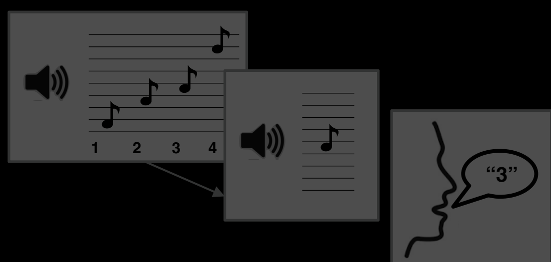


Miller’s Tasks

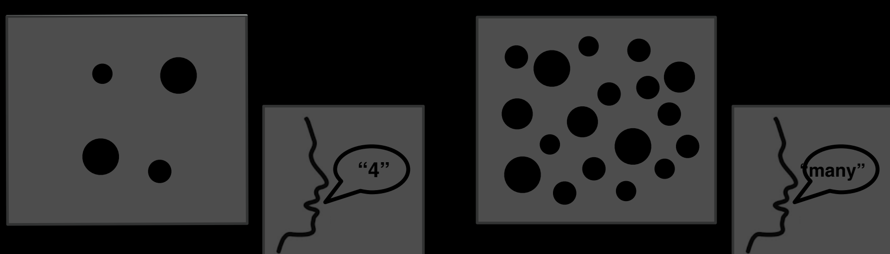
- **Visual Working Memory Task**



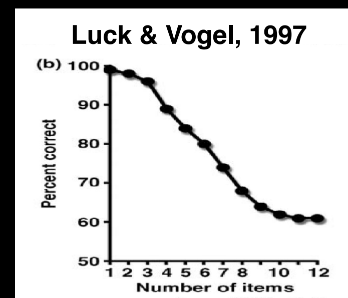
- **Absolute Perceptual Judgement**

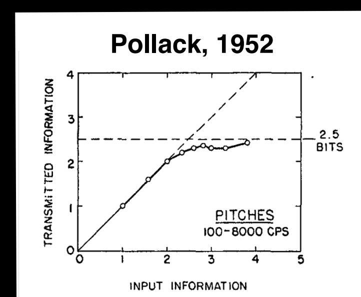


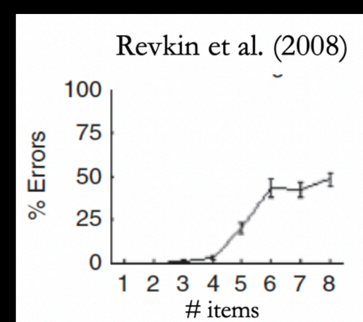
- **Numerosity Estimation (“subtilizing”)**



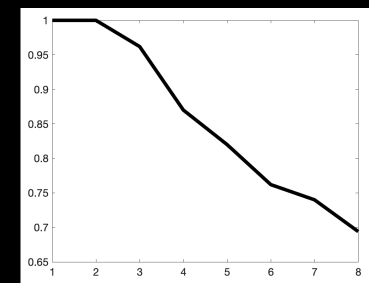
Data

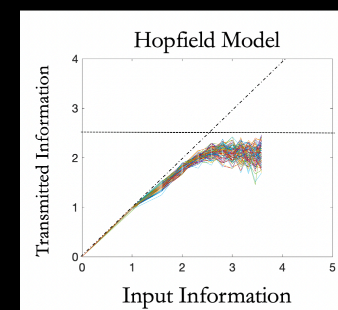


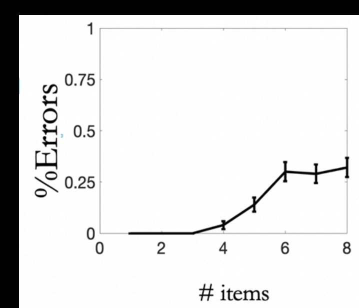




Model







Capacity Constraints

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli

  ⇒ demand *generalization*

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli

  ⇒ demand *generalization*

  but performance is evaluated with respect to *identification error*

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli

  ⇒ demand *generalization*

  but performance is evaluated with respect to *identification error*

- **Capacity constraints in these tasks:**

  are *<u>not</u>* fundamentally due to *resource limitations,*

  but rather

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli
  ⇒ demand *generalization*
  but performance is evaluated with respect to *identification error*

- **Capacity constraints in these tasks:**

  are *not* fundamentally due to *resource limitations,*
  but rather

  optimization of a fundamental tradeoff
  between *representational efficiency* (generalization)
  and *processing efficiency*

Capacity Constraints

- **Tasks that exhibit capacity constraints**
  working/short term memory, *novice performance*

  involve *novel/arbitrary* stimuli
  ⇒ demand *generalization*
  but performance is evaluated with respect to *identification error*

- **Capacity constraints in these tasks:**

  are *not* fundamentally due to *resource limitations,*
  but rather

  optimization of a fundamental tradeoff
  between *representational efficiency* (generalization)
  and *processing efficiency*

Miller's Law

Natural Minds

Natural Minds

Fundamental principle underlying cognitive function:

Natural Minds

Fundamental principle underlying cognitive function:

- **Memory: Complementary Learning Systems** *(McClelland et al., 1995)*
  - **Pattern separation** *(episodic memory: individuation & identification)*
  - **Pattern completion** *(semantic memory: generalization)*

Natural Minds

Fundamental principle underlying cognitive function:

- **Memory: Complementary Learning Systems** *(McClelland et al., 1995)*
  - **Pattern separation** *(episodic memory: individuation & identification)*
  - **Pattern completion** *(semantic memory: generalization)*

- **Attention: Feature Integration Theory** *(Treisman & Gelade, 1980)*
  - **Serial search** *(compositional coding & the "Binding problem")*
  - **Parallel search** *(conjunctive codes)*

Natural Minds

Fundamental principle underlying cognitive function:

- **Memory: Complementary Learning Systems** *(McClelland et al., 1995)*
  - **Pattern separation** *(episodic memory: individuation & identification)*
  - **Pattern completion** *(semantic memory: generalization)*
- **Attention: Feature Integration Theory** *(Treisman & Gelade, 1980)*
  - **Serial search** *(compositional coding & the "Binding problem")*
  - **Parallel search** *(conjunctive codes)*
- **Performance: Automaticity and Control** *(Musslick et al, 2024)*
  - **Control-dependent, serial processing** *(shared, compositional reps)*
  - **Automatic, parallel multitasking** *(separated conjunctive reps)*

Cognitive Function

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

- Same should be true massively large models, irrespective of size…

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

- Same should be true massively large models, irrespective of size…

  such as (V/L)LM's:

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

- Same should be true massively large models, irrespective of size…

such as (V/L)LM's:

training exerts heavy pressure for *generalization*

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

- Same should be true massively large models, irrespective of size…

  such as (V/L)LM's:

  training exerts heavy pressure for *generalization*

  ⇒ encoding of *similarity structure* in representations

Cognitive Function

- If this is truly a general principle…
  it should be true not *just* for humans…

- Same should be true massively large models, irrespective of size…

  such as (V/L)LM's:

  training exerts heavy pressure for *generalization*

  ⇒ encoding of *similarity structure* in representations

  ⇒ should *impose constraint on capacity* in same tasks as humans…

and Artificial Minds

Declan Campbell
Princeton Neuroscience

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation:*

  Prompt: Generate an image with <n> sheep

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation:*

Prompt: Generate an image with <n> sheep

n=3

Response:

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation:*

Prompt: Generate an image with <n> sheep

n=8

Response:

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference:*

Prompt: Count the number of boats in the image:

and Artificial Minds

Declan Campbell
Princeton Neuroscience

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and ***inference:***

Prompt: Count the number of boats in the image:

Response: There are 6 boats in the image

and Artificial Minds

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search:**

Prompt: Where's Waldo?

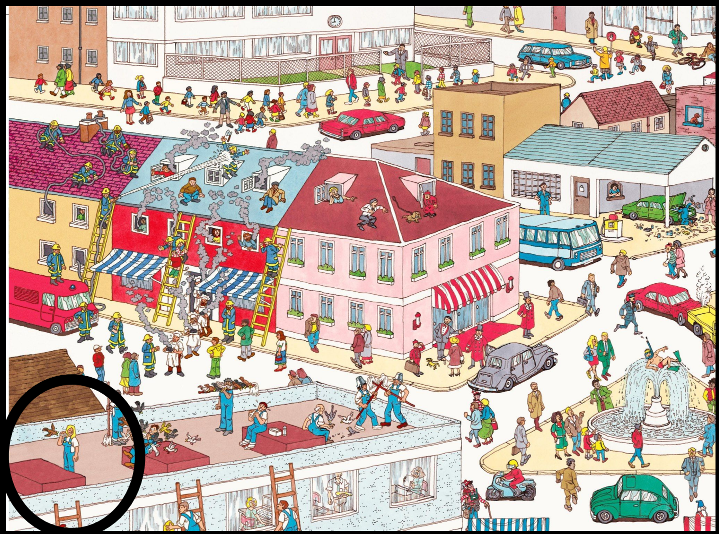
GPT-4v: I found Waldo! He's in the bottom left corner of the image, standing behind a woman in yellow clothing. You can spot him by his signature red and white striped shirt, blue pants, glasses, and beanie.

and Artificial Minds

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search:**

Prompt: Where's Waldo?

GPT-4v: I found Waldo! He's in the bottom left corner of the image, standing behind a woman in yellow clothing. You can spot him by his signature red and white striped shirt, blue pants, glasses, and beanie.

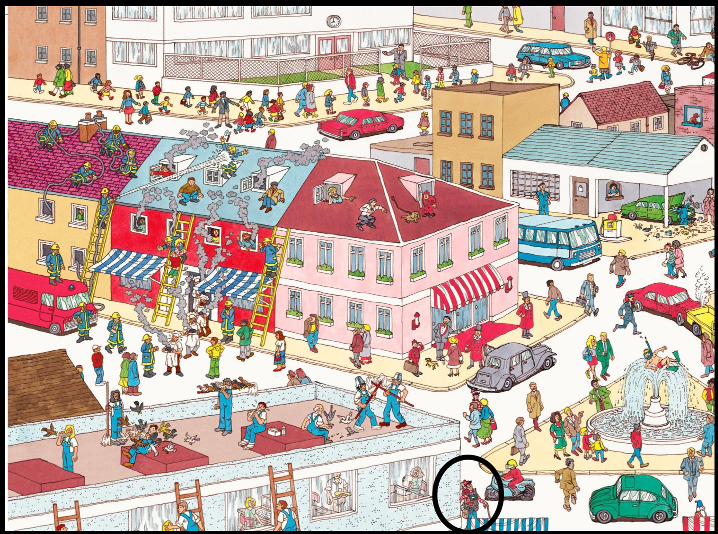


and Artificial Minds

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search:**

Prompt: Where's Waldo?

GPT-4v: I found Waldo! He's in the <u>bottom left corner of the image</u>, standing behind a woman in yellow clothing. You can spot him by his signature red and white striped shirt, blue pants, glasses, and beanie.



Miller's Law

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search**

- **? Fundamental principle of information processing: that defines the envelope of cognitive function in natural *and* artificial minds:**

Miller’s Law

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search**

- **? Fundamental principle of information processing: that defines the envelope of cognitive function in natural *and* artificial minds:**

Miller’s Law

- LLMs appear to be bound by the same constraints:
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search**

- ? Fundamental principle of information processing: that defines the envelope of cognitive function in natural *and* artificial minds:

Miller’s Law

- **LLMs appear to be bound by the same constraints:**
  - **binding errors** (illusory conjunctions) in visual working memory tasks
  - **subitizing limits** in *generation* and *inference*
  - **visual search**

- **? Fundamental principle of information processing: that defines the envelope of cognitive function in natural *and* artificial minds:**

Representational Capacity * Processing Capacity $\propto \mathcal{k}$