Information Generation and the Consciousness Prior

LSE – September 2022

Manolo Martínez

Universitat de Barcelona – Logos – BIAP

Introduction

Information-Processing Theories of Consciousness Then

- Information-processing theories used to mean cognitive-architecture theories

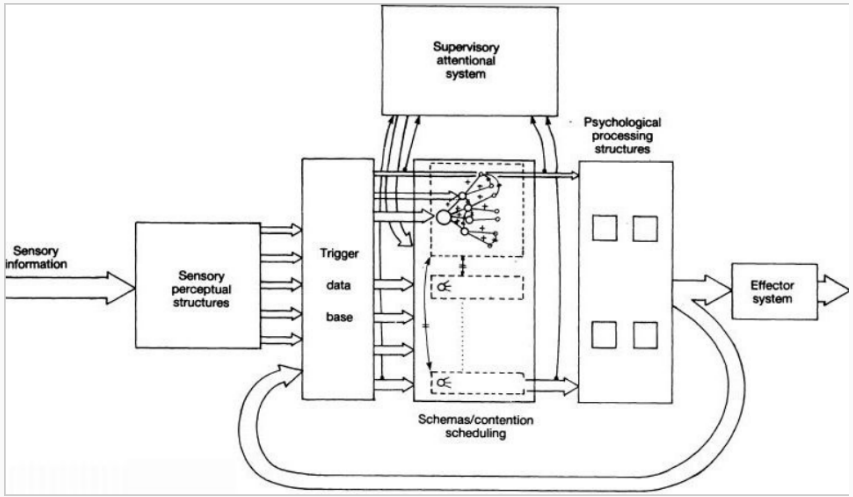


Figure 1: From Norman and Shallice 1980

Information-Processing Theories of Consciousness Then

- Information-processing theories used to mean cognitive-architecture theories

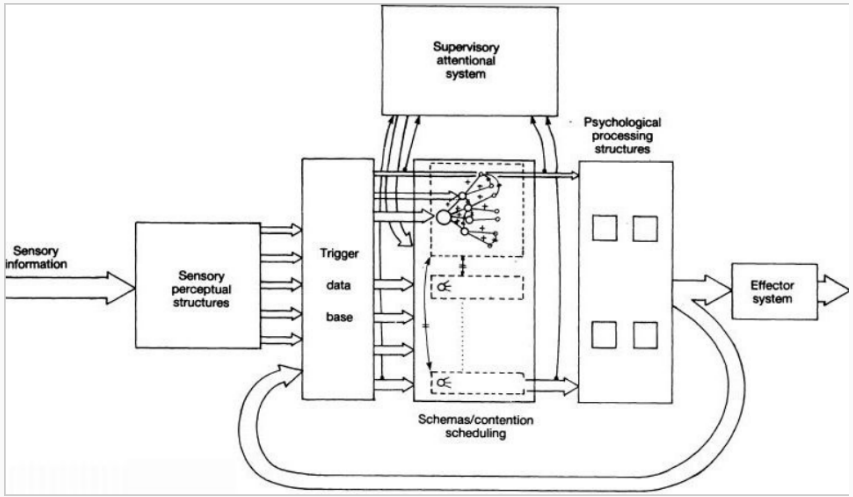


Figure 1: From Norman and Shallice 1980

- "Connections, constraints, subsystems" (Shallice 1988)

Information-Processing Theories of Consciousness Then

- Hardcastle (1995) claims that "information processing theories ... maintain that consciousness is a centralized processor that we use when processing novel or complex stimuli" (p. 89)

Information-Processing Theories of Consciousness Then

- Hardcastle (1995) claims that "information processing theories ... maintain that consciousness is a centralized processor that we use when processing novel or complex stimuli" (p. 89)
- Global workspace accounts [Baars, Dehaene] fall under this general characterization.

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)
    - Conscious information needs to be somehow *integrated*

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)
  - Conscious information needs to be somehow *integrated*
- "Kolmogorov Theory" (Ruffini 2017)

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)
    - Conscious information needs to be somehow *integrated*
- "Kolmogorov Theory" (Ruffini 2017)
    - Conscious information needs to be somehow *compressed*

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)
    - Conscious information needs to be somehow *integrated*
- "Kolmogorov Theory" (Ruffini 2017)
    - Conscious information needs to be somehow *compressed*
- Information Generation (Kanai et al. 2019; Wiese 2020)

Information-Processing Theories of Consciousness Now

- A shift towards attemps to characterize the *informational signature* of consciousness (Wibral et al. 2017)
- IIT (Oizumi, Albantakis, and Tononi 2014; Tegmark 2016)
  - Conscious information needs to be somehow *integrated*
- "Kolmogorov Theory" (Ruffini 2017)
  - Conscious information needs to be somehow *compressed*
- Information Generation (Kanai et al. 2019; Wiese 2020)
  - Consciousness somehow *generates* information

Plan for Today

- Clarify the notion of information generation

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds
  - Certainly not as easy as Kanai et al. make it out to be

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds
  - Certainly not as easy as Kanai et al. make it out to be
- Connect this idea with others usually floated in the debate on consciousness

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds
  - Certainly not as easy as Kanai et al. make it out to be
- Connect this idea with others usually floated in the debate on consciousness
  - The limited capacity of consciousness

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds
  - Certainly not as easy as Kanai et al. make it out to be
- Connect this idea with others usually floated in the debate on consciousness
  - The limited capacity of consciousness
  - Its reliance on short-term memory

Plan for Today

- Clarify the notion of information generation
  - It's not as easy to generate information as it sounds
  - Certainly not as easy as Kanai et al. make it out to be
- Connect this idea with others usually floated in the debate on consciousness
  - The limited capacity of consciousness
  - Its reliance on short-term memory
- And one less usually floated: Bengio's *consciousness as a prior* (2019)

The Function of Consciousness

Cognition

- I'll just assume that if consciousness is for anything, it is for cognition

Cognition

- I'll just assume that if consciousness is for anything, it is for cognition
  - It is there to help with cognitive function

Cognition

- I'll just assume that if consciousness is for anything, it is for cognition
  - It is there to help with cognitive function
- This is compatible with 4E approaches to cognition

Cognition

- I'll just assume that if consciousness is for anything, it is for cognition
  - It is there to help with cognitive function
- This is compatible with 4E approaches to cognition
  - Cognition as the production of adaptive behavior (Barack and Krakauer 2021, 359)

Cognition

- I'll just assume that if consciousness is for anything, it is for cognition
  - It is there to help with cognitive function
- This is compatible with 4E approaches to cognition
  - Cognition as the production of adaptive behavior (Barack and Krakauer 2021, 359)
  - Perceptually guided action (Varela 2017, 173)

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body
2. A variable B capturing the behavioral output of an agent

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body
2. A variable B capturing the behavioral output of an agent
3. A distortion measure (loss function, objective function) connecting C and B

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body
2. A variable B capturing the behavioral output of an agent
3. A distortion measure (loss function, objective function) connecting C and B

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body
2. A variable B capturing the behavioral output of an agent
3. A distortion measure (loss function, objective function) connecting C and B

$$d \colon C \times B \to \mathbb{R}^+$$

An Ecumenical Model of Cognition

1. A variable C capturing the state of the world including the agent's body
2. A variable B capturing the behavioral output of an agent
3. A distortion measure (loss function, objective function) connecting C and B

$$d \colon C \times B \to \mathbb{R}^+$$

Main Model: The transformation of one variable, C, into another, B, in a way that minimizes $d(C, B)$.

Why Consciousness?

- Why is consciousness necessary, or useful, for main-model tasks?

Why Consciousness?

- Why is consciousness necessary, or useful, for main-model tasks?
  - It is unnecessary for many such tasks (Dehaene 2001)

Why Consciousness?

- Why is consciousness necessary, or useful, for main-model tasks?
  - It is unnecessary for many such tasks (Dehaene 2001)
- More generally: the d-optimal mapping between C and B can be given by a lookup table

Why Consciousness?

- Why is consciousness necessary, or useful, for main-model tasks?
  - It is unnecessary for many such tasks (Dehaene 2001)
- More generally: the d-optimal mapping between C and B can be given by a lookup table
  - So with unlimited-capacity channels, memory, and computational resources there's no need for any consciousness tricks

Why Consciousness?

- Why is consciousness necessary, or useful, for main-model tasks?
  - It is unnecessary for many such tasks (Dehaene 2001)
- More generally: the d-optimal mapping between C and B can be given by a lookup table
  - So with unlimited-capacity channels, memory, and computational resources there's no need for any consciousness tricks
- Consciousness must be related to circumventing computational-complexity and capacity limitations

Figure 2: Drawn by Stable Diffusion

Information Generation

Kanai et al.

- A core function of consciousness is the generation of information

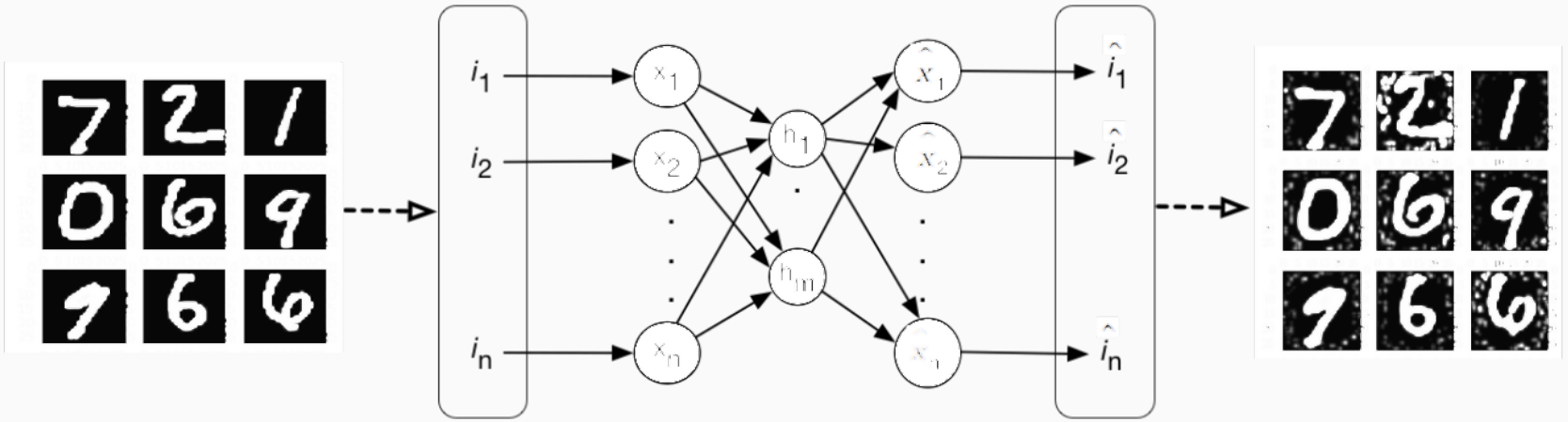


Figure 3: pclub.in

Kanai et al.

- A core function of consciousness is the generation of information
- Their intuitive model is autoencoders

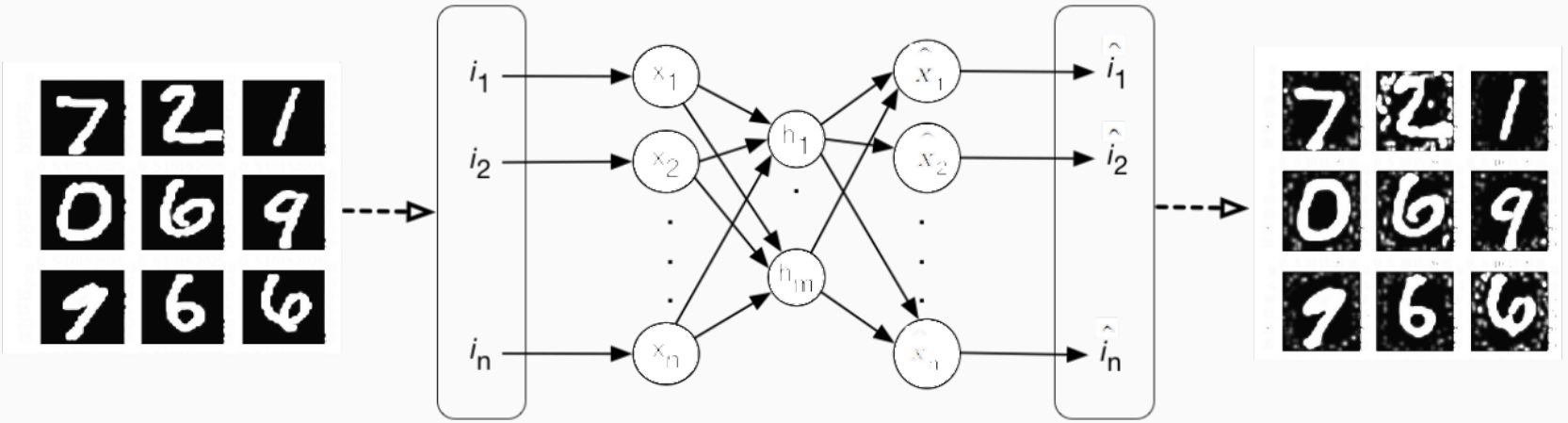
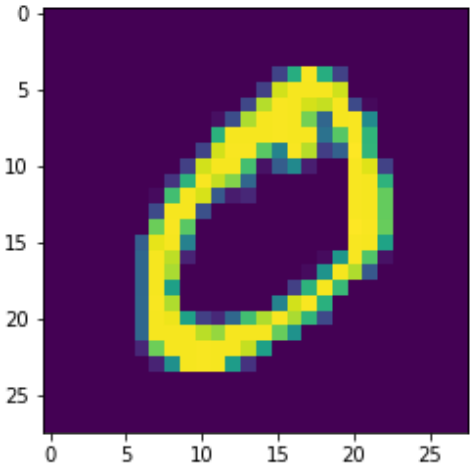


Figure 3: pclub.in

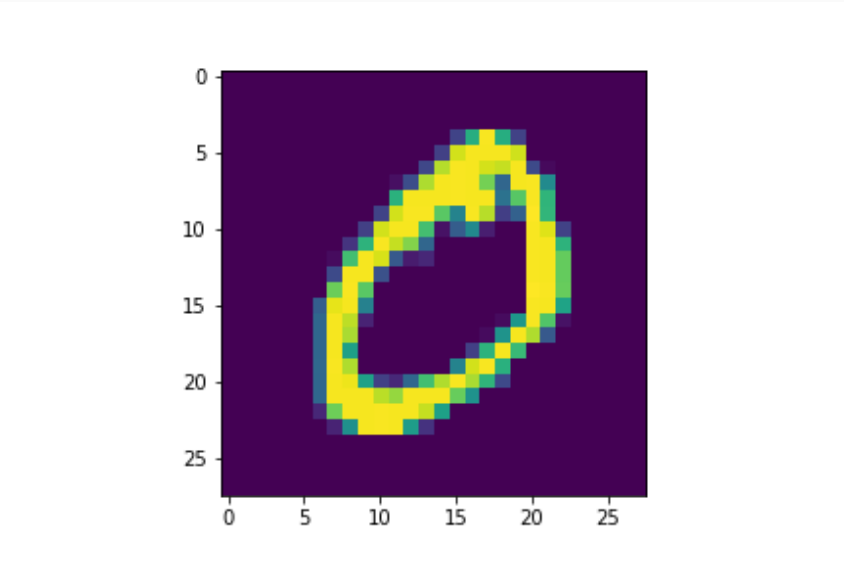
MNIST and VAEs

- MNIST images have $28 \times 28 = 784$ pixels (that is to say, dimensions)



MNIST and VAEs

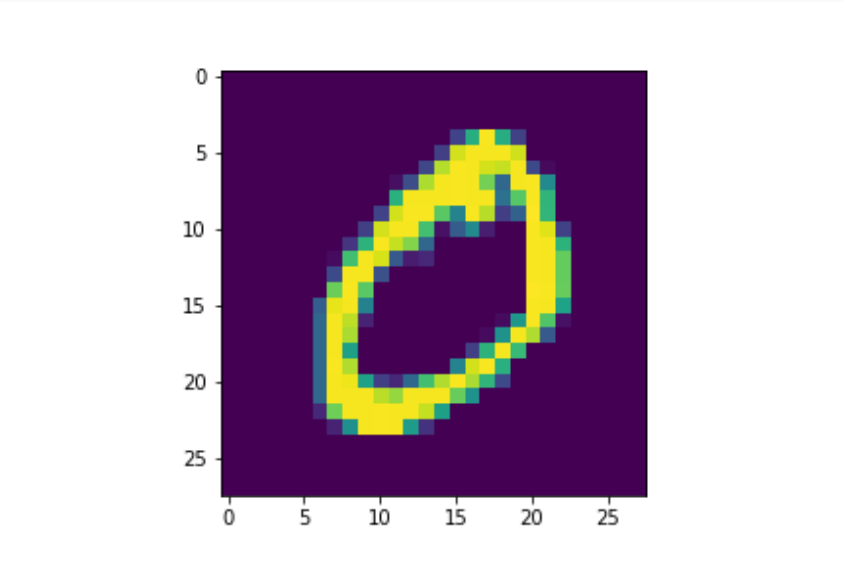
- MNIST images have $28 \times 28 = 784$ pixels (that is to say, dimensions)



- But in fact the "intrinsic dimensionality" of MNIST samples is much lower

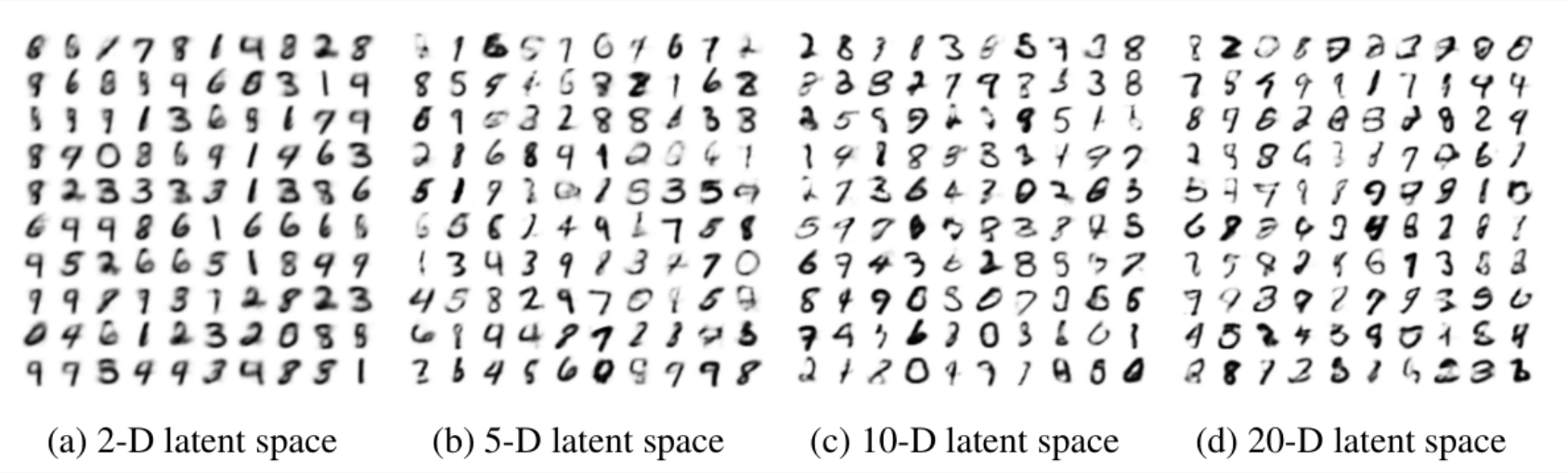
MNIST and VAEs

- MNIST images have $28 \times 28 = 784$ pixels (that is to say, dimensions)



- But in fact the "intrinsic dimensionality" of MNIST samples is much lower
  - Lower-dimensional latent spaces suffice

Decoding as Information Generation



(a) 2-D latent space (b) 5-D latent space (c) 10-D latent space (d) 20-D latent space

Figure 4: From Kingma and Welling (2014)

- According to Kanai et al., the decoder generates information when it reconstructs MNIST digits from points in a low-dimensional latent space.

The CSI fallacy

Figure 5: knowyourmeme

- But this is not information generation

The CSI fallacy

Figure 5: knowyourmeme

- But this is not information generation
  - Either the information was in the original image, or it wasn't

The CSI fallacy

Figure 5: knowyourmeme

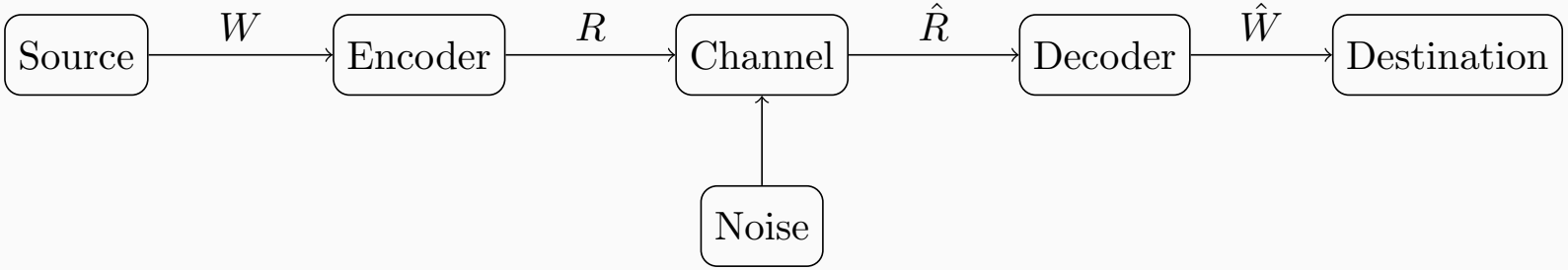
- But this is not information generation
  - Either the information was in the original image, or it wasn't
  - If it wasn't, we are in "Zoom and Enhance!" meme territory

The CSI fallacy

Figure 5: knowyourmeme

- But this is not information generation
  - Either the information was in the original image, or it wasn't
  - If it wasn't, we are in "Zoom and Enhance!" meme territory
  - If it was, nothing is generated

The Data-Processing Inequality



$$I(\hat{W}; W) \leq I(R; W)$$

(Cover and Thomas 2006, 34)

Synergies Across Time

A Better Example of Information Generation

- The closer we can get to actual information generation is *synergistic informational contributions* (Williams and Beer 2010; Wibral et al. 2017)

A Better Example of Information Generation

- The closer we can get to actual information generation is *synergistic informational contributions* (Williams and Beer 2010; Wibral et al. 2017)
  - Two or more variables *jointly* carrying more information about a variable of interest than the sum of their individual contributions.

A Better Example of Information Generation

- The closer we can get to actual information generation is *synergistic informational contributions* (Williams and Beer 2010; Wibral et al. 2017)
  - Two or more variables *jointly* carrying more information about a variable of interest than the sum of their individual contributions.
  - (More or less)

An Example of Synergistic Variables

A	B	C	Pr
0	0	0	.25
0	1	1	.25
1	0	1	.25
1	1	0	.25

- A carries no information about C

An Example of Synergistic Variables

A	B	C	Pr
0	0	0	.25
0	1	1	.25
1	0	1	.25
1	1	0	.25

- A carries no information about C
- B carries no information about C

An Example of Synergistic Variables

A	B	C	Pr
0	0	0	.25
0	1	1	.25
1	0	1	.25
1	1	0	.25

- A carries no information about C
- B carries no information about C
- AB is perfectly informative about C

Synergy and Main-Model Tasks

- If we have different sources of information, e.g., encoded in different sensory variables ...

Synergy and Main-Model Tasks

- If we have different sources of information, e.g., encoded in different sensory variables ...
  - This is a version of the "modules" idea

Synergy and Main-Model Tasks

- If we have different sources of information, e.g., encoded in different sensory variables ...
  - This is a version of the "modules" idea
- ... then massive parallel processing (in general) won't do

Synergy and Main-Model Tasks

- If we have different sources of information, e.g., encoded in different sensory variables ...
  - This is a version of the "modules" idea
- ... then massive parallel processing (in general) won't do
  - Synergistic information carried by groups of those variables will be lost

Synergy and Main-Model Tasks

- If we have different sources of information, e.g., encoded in different sensory variables ...
  - This is a version of the "modules" idea
- ... then massive parallel processing (in general) won't do
  - Synergistic information carried by groups of those variables will be lost
- At some point, those variables will need to be put in common

Synergistic Information and Consciousness

- The idea that consciousness has an “integrative function” (Baars 2005)

Synergistic Information and Consciousness

- The idea that consciousness has an "integrative function" (Baars 2005)
  - In GW theories, integration is mostly about outputs, not inputs

Synergistic Information and Consciousness

- The idea that consciousness has an "integrative function" (Baars 2005)
  - In GW theories, integration is mostly about outputs, not inputs
  - More on this later.

Synergistic Information and Consciousness

- The idea that consciousness has an "integrative function" (Baars 2005)
  - In GW theories, integration is mostly about outputs, not inputs
  - More on this later.
- The main IIT idea

Synergistic Information and Consciousness

- The idea that consciousness has an "integrative function" (Baars 2005)
  - In GW theories, integration is mostly about outputs, not inputs
  - More on this later.
- The main IIT idea
  - ϕ is just a measure of synergistic information (Griffith et al. 2014)

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*
- Diachronic synergy is more interesting:

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*
- Diachronic synergy is more interesting:
  - Think of the world as a Markov process: $W_1 \to W_2 \to \cdots \to W_n$

Synergy, synchronic and diachronic

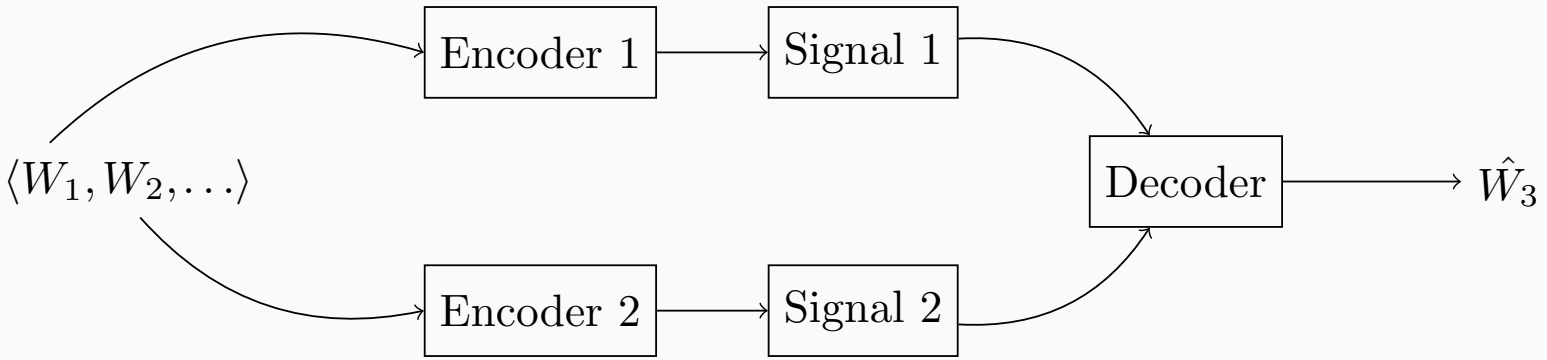
- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*
- Diachronic synergy is more interesting:
  - Think of the world as a Markov process: $W_1 \to W_2 \to \cdots \to W_n$
  - It's entirely possible to have information that is neither in W_i nor in W_j but is synergistically in $W_i W_j$

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*
- Diachronic synergy is more interesting:
  - Think of the world as a Markov process: $W_1 \to W_2 \to \cdots \to W_n$
  - It's entirely possible to have information that is neither in W_i nor in W_j but is synergistically in $W_i W_j$

Synergy, synchronic and diachronic

- If you squint hard enough, synchronic synergy looks a bit like information generation
  - Information in the decoded message that is in no sensory input
  - But it always is in sensory input *as a whole*
- Diachronic synergy is more interesting:
  - Think of the world as a Markov process: $W_1 \to W_2 \to \cdots \to W_n$
  - It's entirely possible to have information that is neither in W_i nor in W_j but is synergistically in $W_i W_j$



The Consciousness Prior

Taking Stock

- Information generation is not a thing

Taking Stock

- Information generation is not a thing
- But synergistic relations get as close as possible

Taking Stock

- Information generation is not a thing
- But synergistic relations get as close as possible
    - They are a common motif between IIT, GWT, and information generation accounts

Taking Stock

- Information generation is not a thing
- But synergistic relations get as close as possible
    - They are a common motif between IIT, GWT, and information generation accounts
- ... And synergistic relations in time perhaps closest of all

Short-Term Memory

- Short-term memory is usually considered as central to consciousness (Ramachandran and Hirstein 1997; Dehaene 2001)

Short-Term Memory

- Short-term memory is usually considered as central to consciousness (Ramachandran and Hirstein 1997; Dehaene 2001)
  - One additional reason is that information about W_i must be kept around so as to combine it with subsequent W_j

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that
  1. Task-relevant information will often be carried synergistically by different modules

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that
  1. Task-relevant information will often be carried synergistically by different modules
  2. It is possible to construct one-size-fits-all packages of information that can be widely broadcast (hence its being low-capacity – and the reliance on attention)

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that
  1. Task-relevant information will often be carried synergistically by different modules
  2. It is possible to construct one-size-fits-all packages of information that can be widely broadcast (hence its being low-capacity – and the reliance on attention)
- Its reliance on short-term memory is (at least partly) the expectation that

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that
  1. Task-relevant information will often be carried synergistically by different modules
  2. It is possible to construct one-size-fits-all packages of information that can be widely broadcast (hence its being low-capacity – and the reliance on attention)
- Its reliance on short-term memory is (at least partly) the expectation that
  1. There will be synergistic connections between past and present (hence the need for memory)

The Consciousness Prior

- Bengio (2019) thinks of the GW architecture as a *prior*
  - What he means by this is that consciousness is optimized for a way the world is expected to be
- Its "integrative function" is the expectation that
  1. Task-relevant information will often be carried synergistically by different modules
  2. It is possible to construct one-size-fits-all packages of information that can be widely broadcast (hence its being low-capacity – and the reliance on attention)
- Its reliance on short-term memory is (at least partly) the expectation that
  1. There will be synergistic connections between past and present (hence the need for memory)
  2. Those synergistic connections mostly decay after a while (hence its being short term)

Conclusions

Conclusions

- It's useful to think of consciousness in terms of its informational signature

Conclusions

- It's useful to think of consciousness in terms of its informational signature
  - Abstracting away from mechanism

Conclusions

- It's useful to think of consciousness in terms of its informational signature
  - Abstracting away from mechanism
  - Consciousness is a "systems" response to certain information-processing needs

Conclusions

- Information generation does not exist

Conclusions

- Information generation does not exist
- Synergistic connections is the next best thing

Conclusions

- Information generation does not exist
- Synergistic connections is the next best thing
    - Exploiting them is part of what global workspaces + attention do

Conclusions

- Information generation does not exist
- Synergistic connections is the next best thing
    - Exploiting them is part of what global workspaces + attention do
- They are not just synchronic but diachronic

Conclusions

- Information generation does not exist
- Synergistic connections is the next best thing
    - Exploiting them is part of what global workspaces + attention do
- They are not just synchronic but diachronic
    - Exploiting *those* is part of what short-term memory does

Conclusions

- Information generation does not exist
- Synergistic connections is the next best thing
    - Exploiting them is part of what global workspaces + attention do
- They are not just synchronic but diachronic
    - Exploiting *those* is part of what short-term memory does
    - It is a component of the "consciousness prior" that the bulk of these connections fade out rather quickly (hence short term)

Thanks!

manolomartinez.net

mail@manolomartinez.net

manolomartinez

References

Baars, Bernard J. 2005. "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience." In *Progress in Brain Research*, 150:45–53. Elsevier. https://doi.org/10.1016/S0079-6123(05)50004-9.

Barack, David L., and John W. Krakauer. 2021. "Two Views on the Cognitive Brain." *Nature Reviews Neuroscience* 22 (6): 359–71. https://doi.org/10.1038/s41583-021-00448-6.

Bengio, Yoshua. 2019. "The Consciousness Prior." arXiv. https://doi.org/10.48550/arXiv.1709.08568.

Cover, T. M., and J. A. Thomas. 2006. *Elements of Information Theory*. New York: Wiley.

Dehaene, S. 2001. "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework." *Cognition* 79 (1-2): 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2.

Griffith, Virgil, Edwin KP Chong, Ryan G. James, Christopher J.