KEITH FRANKISH (2021)

# TECHNOLOGY & THE HUMAN MINDS

*Casted into powerpoint*

*by Anna Strasser*

1. introducing the dual-process approach
2. re-interpretation of dual-process theory
3. implications for projects of cognitive enhancement & AI
4. assessing the risks of those projects

# INTRODUCTION

## OVER THE LAST 40 YEARS, MANY PSYCHOLOGISTS HAVE COME TO ADOPT SOME FORM OF *DUAL-PROCESS* THEORY

human cognition is supported by two distinct types of processing which can yield different and sometimes conflicting results

- fast, automatic, unconscious
  → INTUITION *(spontaneous responses that just feel right)*

- slower, controlled, conscious
  → REFLECTION *(more considered responses for which one can give some explicit justification)*

two distinct types of processing associated with two brain systems

System 1

- evolutionarily ancient and largely shared with other animals

System 2

- more recent and distinctively human.

→ obvious implications for cognitive enhancement and artificial intelligence (AI)

projects of enhancing & replicating human intelligence will each also assume a dual aspect

(Evans 2010)
**having 2 minds**
intuitive *old mind* &
reflective *new mind*

**BUT**

Dual-process views face some problems, and it is hard to see how System 2 could be modelled artificially.

→ reinterpretation of dual-process theory picturing the two minds as levels of organization rather than distinct systems

- new mind should be seen as a virtual one (*not as a brain system*) formed by culturally transmitted habits which restructure the activities of the old mind
- resolving some of the problems for dual-process theory & making the project of artificially creating a System 2 mind somewhat more tractable

# Dual Processes

**from the 1970s onwards**

*grew out of experimental work in cognitive & social psychology*
(Frankish 2010)

**in the late 1980s & 1990s**

*published*

(e.g., Chaiken & Trope 1999; Chen & Chaiken 1999; Epstein 1994; Evans 1989; Evans & Over 1996; Petty & Cacioppo 1986; Sloman 1996; Stanovich 1999; Stanovich & West 2000)

**over the next decade**

*brought to a wider audience*

(Evans 2010; Kahneman 2011; Stanovich 2004).

agreeing on the fundamentals

- ~~two types of processing ('thinking')~~ involved in human reasoning, decision making, social cognition

differences as to how the two processes are related

1. operating independently & competing for control of behavior
2. *default-interventionist* model (Evans 2006; Kahneman 2011):
   Type 1 supply rapid default responses, which can be intervened upon & overridden by Type 2
   Type 1 processes trigger Type 2 processing & select information for it to use

| Type 1 processing | Type 2 processing |
|---|---|
| typically fast, automatic, effortless, non-conscious, associative, parallel, high-capacity, undemanding of working memory | typically slow, controlled, effortful, conscious, rule-governed, serial, low capacity, demanding of working memory |
| highly contextualized<br>draws on implicit knowledge acquired from past experience | more abstract<br>draws on explicit knowledge & learned rules of inference |
| delivers responses that may be adaptive in real-world settings but often deviate from rational norms<br>- manifesting cognitive biases, stereotype effects, emotional influences | more likely to deliver responses in line with normative principles |

linked to hypothetical thinking
enabling 'secondary' representations
- decoupled from the world & do not directly affect behaviour

propensity to use Type 2 processing
- high individual variability
- correlated with measures of general intelligence

# A BROADER ARCHITECTURAL BASIS FOR THE TWO TYPES OF PROCESSING

## System 1

- composed of multiple subsystems
  - many evolutionarily ancient
  - operating in a Type 1 way (e.g., Stanovich 2004)
- perceptual, motivational, emotional systems, learning & conceptual systems
  - perhaps specialized for particular tasks
    *(navigation, foraging, social cognition, theory of mind, language)*
- procedures for learned skills practiced to automaticity *(reading & driving)**

## System 2

- single, low-capacity system which can manipulate explicit representations in working memory
- flexible
- responsive to instructions
- uniquely human

mass of evidence for the dual-process picture from three independent sources (Evans & Stanovich 2013)

### 1. response patterns in reasoning & decision-making tasks

- *two kinds of answers*
  1. intuitively plausible but normatively incorrect
  2. less obvious but correct
- *experimental manipulations:*
  - time pressure → increased production of intuitive answers (rather than random responding)
  - clear task instructions → normatively correct one

two different mechanisms in play,
one fast & intuitive, the other slow & reflective

### 2. individual differences

- positive correlation between tendency to give the normative responses on reasoning tasks & general intelligence

hypothesis:

- higher general intelligence → more working memory capacity → greater capacity to engage in and sustain Type 2 processing and to override intuitive Type 1 responses

### 3. neuroscientific evidence

- Imaging studies indicate that different neural structures are involved in the production of responses associated with each type of processing

Type 2 responses typically following activation of prefrontal & frontal cortical regions that are not involved in Type 1 responding.

* The basic dual-system framework is compatible with a spectrum of views as to the nature of the evolved components of System 1, from ones which posit multiple domain-specific modules (e.g., Carruthers 2006) to ones which hold that learning is domain-general and that specialized systems are *cognitive gadgets* installed by cultural processes during individual development (Heyes 2018).

This dual-system view has a common-sense appeal, has been tacitly acknowledged for centuries (Frankish & Evans 2009).

- Descartes
  - identifying the mind with the conscious mind, understood as an immaterial substance that is the arena of pure thought.
  - also recognized that much human & animal behavior occurs without conscious thought and must be supported by complex nonconscious mechanisms of some kind.

- in the nineteenth century
  - gradual acceptance that these processes were genuinely mental, involving non-conscious perceptions & thoughts, operating independently of the conscious mind

- more recently
  - development of the computational theory of mind & modern cognitive science

  → non-conscious processes increasingly took center-stage in the explanation of human behavior, with the conscious mind sometimes being demoted to the role of a rationalizer (Wegner 2002)

history of AI

implicit acknowledgment of the two-systems distinction

- Early AI researchers: focus on System 2
  - focused on abstract reasoning and decision-making, which they sought to model in computational terms, with the aim of creating artificial general intelligence

- Lack of success: focus on System 1
  - many researchers turn to a bottom-up approach, seeking to create embodied, robotic systems with specific behavioral competences (e.g., Brooks 1991; Steels and Brooks 1995)

# PROBLEMS

critics

Gigerenzer 2010; Keren & Schul 2009; Kruglanski & Gigerenzer 2011; Melnikoff & Bargh 2018; Osman 2004, 2018

common objection

- it is highly unlikely that
  - various features ascribed to each process *(fast vs slow, automatic vs controlled, non-conscious vs conscious, etc.)* align so neatly
  - one can exclude crossover processes that are (*fast but controlled*)
  - intuitive processing is always biased
  - reflective processing is always normatively rational

Dual-process theorists

- clarifying the scope of their claims (e.g., Evans & Stanovich 2013; Pennycook et al. 2018)
- features ascribed to each process are not all defining ones & the core distinction can be drawn more simply

Evans & Stanovich

System 1: autonomy (lack of attentional control) | System 2: use of working memory & support for decoupled representations as those of (Evans & Stanovich 2013; Stanovich & Toplak 2012)

- other features just merely typical correlates of these defining features
  - Type 2 processing is typically slow & serial because it loads on working memory
  - allows for considerable variation in the mode of Type 2 thinking (Stanovich 2009a, 2011)
- wrong to think (even though often observed in experimental settings)
  - Type 2 processes always deliver normatively correct responses
  - all cognitive errors are due to Type 1 processes.
- no universal conflict between the two kinds of processing

more specific worries about the relation between the two systems

→ prompted proposals for the revision or refinement of the framework, though without undermining the case for a qualitative distinction along the general lines proposed (De Neys 2018)

# A GENERAL PROBLEM

concerns Type 2 processing

## WHAT EXACTLY IS THE MECHANISM BY WHICH THIS PROCESSING OPERATES?

- Theorists identify various components of this device (*working memory, explicit decoupled representations, executive control processes*) but these do not in themselves amount to a reasoning system.

**What is the engine that manipulates the explicit representations in working memory, in accordance with rules of inference or other procedures?**

Dual-system theorists are strangely silent on this.

related evolutionary worry

If System 2 is an extraordinarily powerful self-contained device, how & why did such a system evolve?

- evolution seems to have involved the addition of new specialist subsystems, such as ones for language, mindreading, and social cognition, and the enhancement of existing ones,

BUT WHY ASSUME A COMPLETELY NEW GENERAL-PURPOSE REASONING SYSTEM? (Carruthers 2006)

What evolutionary pressures could there have been for the development of such a system?

- Having a capacity for flexible, abstract, rule-governed deliberation is advantageous in the modern world (a world that is largely the creation of our human minds), but it is hard to see why it would have been required in the ancestral environment in which our species evolved.
- Cognitive flexibility is certainly useful, but building in general intelligence seems like a massively overengineered solution to any specific environmental challenges our ancestors might have faced.

# TYPE 2 THINKING AS AN ACTIVITY

**REINTERPRETATION OF DUAL-PROCESS THEORY**

key idea : some thinking = intentional activity, something we *do*

- under voluntary control, responding to our beliefs & desires
  we *want* to | we enjoy them | we believe they will further some goal we have
- reasons need not be consciously entertained

UNREFLECTIVE INTENTIONAL BEHAVIOR:

- without giving much conscious thought to the reasons for our actions
- directed to our goals & guided by our beliefs

DEFINING FEATURE

- Type 2 thinking
  - involves performing intentional actions
- Type 1 processing
  - wholly automatic process
  - occurs without our needing to do anything

How could reasoning be an intentional activity?

*solving a long division problem with pencil & paper*

- controlled, conscious procedure = series of actions with the goal of solving the problem
  1. writing down various numerals in certain locations
  2. do a series of simpler calculations → System 1:comes to us intuitively, automatic Type 1 processes, neither controlled nor conscious

DELIBERATIVE *MASTICATION*

AUTOSTIMULATION (Dennett 1991, Ch. 7)

in creating & manipulating external symbols, we are cognitively stimulating ourselves

providing new inputs to our Type 1 mental processes

- perceptual systems detect & interpret the symbols we create
- conceptual, emotional, and motivational systems get to work on the problem of how to respond
- compete for control of motor systems, leading to a further action, which forms the next step in the sequence

*autostimulatory processes*

- drawings & diagrams to help us solve problems & evaluate options
- making sketches to experiment with designs for a garden

→ our autonomous mental processes generate an evaluative response

# SPEECH

## THE MOST POWERFUL MEANS OF AUTOSTIMULATION

By talking to ourselves we can work our way through a tricky problem

- question ourselves ('Where did I leave the remote?')
- guide ourselves ('That's the earth pin, so this must be the live')
- prompt ourselves ('It begins with a T')
- encourage ourselves ('You can do it!')
- chide ourselves ('Focus!')

utterances are intentional actions

performed with the goal of solving our current problem.

- processed like other utterances & interpreted as requiring some response
  - ➢ Type 1 processes get to work on the task
  - ➢ with luck, generate a further utterance or other action which either solves our problem or takes us a step closer to a solution

Sometimes we conduct a dialogue with ourselves, posing questions and answering them as a way of thinking through the options.

- create extended arguments by moving from one utterance to another in accordance with simple inferential principles we have been taught or have picked up in the course of debate with others
- we combine a variety of techniques to explore a problem space, using utterances as cognitive stepping stones

EXCELLENT MEDIUM FOR FLEXIBLE, REFLECTIVE THINKING,

HAVING AN OPEN-ENDED REPRESENTATIONAL CAPACITY & A SYNTACTIC STRUCTURE THAT FACILITATES LOGICAL INFERENCE

INSTEAD OF PRODUCING OVERT SYMBOLS, SKETCHES, UTTERANCES, WE CAN INTENTIONALLY CREATE MENTAL IMAGERY

- inner speech
  - mentally rehearsing the action of saying the words in question (which generates sensory representations of hearing them),
  - intentional direction of attention in order to stimulate sensory activity associated with relevant stimuli or with episodic memories (Carruthers 2015)

IMAGERY PRODUCED HAS AN AUTOSTIMULATORY EFFECT.

- attention sustains the representations in working memory, resulting in their being made available ('globally broadcast') to all Type 1 subsystems, which process them as they would representations generated by external stimuli

# MENTAL IMAGERY ALLOWS

- internalization of many external problem-solving activities
  - processes of self-questioning, self-guiding, self-prompting, argument construction, inner dialogue can be conducted silently in one's head
- development of a wide range of new problem-solving strategies
  - imagined scenarios serve as proxies for aspects of the world

**Visual imagery, together with imaged utterances, provide the decoupled 'secondary' representations needed for hypothetical thinking.**

## VARIOUS TECHNIQUES OF IMAGISTIC AUTOSTIMULATION

- can be flexibly combined in an exploratory way
- needn't be pre-planned
  - we don't need to know precisely which auto- stimulations to generate in order to solve a problem
  - process of trial & error and may hit many dead ends before we reach a solution
- needn't be completely random
  - we may have picked up useful tricks & developed hunches about what will work, based on past experience

In fact, I believe that access consciousness is the only kind there is and that phenomenal consciousness is illusory (Frankish 2016). But that is another – though related – story.

## CORE DISTINCTION BETWEEN TYPE 1 & TYPE 2 PROCESSING CONCERNS

## THE ROLE OF INTENTIONAL AUTOSTIMULATORY ACTIONS

Type 2 constitutively involve autostimulatory actions | Type 1 does not (not under intentional control, rather autonomous)

**defining characteristics of Type 2 reasoning**

- involves intentional autostimulatory actions
- covert or overt (not restricted to ones that occur 'in the head', using sensory imagery)
- also involves Type 1 processing and is driven by it
- Intentional autostimulation loads on working memory & supports cognitive decoupling since the perceptual or imagistic representations involved are held in working memory and can represent non-actual states of affairs.

→ typical correlated features:

- conscious because representations generated are globally broadcast
- controlled because it is an intentional action
- slow & effortful because it requires controlled attention
- serial because we can perform only one action at a time ….

# A VIRTUAL MIND

## IMPLICATIONS FOR THE EVOLUTION OF THE NEW, 'SYSTEM 2' MIND

predominantly a process of cultural evolution + discovery & transmission of habits of autostimulation.

| NO – because already there | NEW – discovered & learned |
|---|---|
| creation of a new general-purpose reasoning system | natural language* probably developed initially for social purposes |
| completely new neural structures | process of cultural evolution |
| specialist subsystems (*perceptual, conceptual, emotional, motivational*)<br>• constitute the old, System 1 mind<br>• evolved in response to specific adaptive pressures | discovery of habits of autostimulation |
| forms of working memory, attention, episodic memory, executive control<br>• found in other animals (Carruthers 2015, Ch. 8) | transmission of habits of autostimulation |

### PRIVATIZATION & INTERNALIZATION OF CERTAIN SOCIAL PRACTICES

1. *social practices:*
   - cognitively stimulating each other, helping their peers solve problems by offering suggestions, giving advice, asking questions, making sketches
   - practices of public argumentation, setting out arguments in favor of their ideas & plans
2. *privatization:*
   - providing a similar commentary on their own activities and constructing arguments in private
3. *internalization*
4. developing further self-stimulatory tricks using mental imagery

### RELATIVELY MINOR NEURAL ADAPTATIONS TO SUPPORT THE PROCESS

- discovery of autostimulation → huge advantage over peers, creating selectional pressure for neural adaptations favoring the automatic acquisition & elaboration
  - a process known as the Baldwin effect (Dennett 1991)

### LEARNED TECHNIQUES OF INTENTIONAL REASONING

a parallel process occurs in child development
(Diaz & Berk 1992; Vygotsky 1986; Winsler et al. 2009)

1. *social practices:* adults scaffold children's cognitive development by offering guidance, suggestions, instructions
2. *privatization:* children imitate this commentary in self-directed ('private') speech
3. *internalization:* private speech → inner speech

* Speculating about the origins of language is risky; evolution was initially driven by the needs & opportunities of social life, though its co-option for cognitive purposes may have fostered its further development → evolutionary process itself was a combined biological and cultural one (Dennett 2017).

# REINTERPRETATION OF DUAL-PROCESS THEORY

**THERE IS JUST ONE NEURAL SYSTEM – NO SUBSYSTEMS DESIGNED SOLELY TO SUPPORT TYPE 2 THINKING**

collection of 'System 1' subsystems with attentional & executive systems & working memory is sufficient

- compatible with the neuroimaging evidence for dual-process theory
  - not exactly the same subsystems are involved in generating a Type 2 response to a problem
  - Type 2 thinking may bring a wider range of neural resources to bear on the problem & involves engaging executive & working memory systems

→ **'System 2'**

❖ not a neural system

❖ but a new level of organization

  ❖ formed by culturally transmitted habits restructuring the activities of the biological brain

It is a softwired 'virtual machine', like a computer operating system, running on the hardware of the biological brain!

If the old mind is a biological mind, then the new mind is a virtual one

# REINTERPRETATION OF DUAL-PROCESS THEORY

## A SLEIGHT OF HAND?

- How could perceptual and imagistic feedback so radically enhance the problem-solving powers of the brain?

*knowledge that we draw on in Type 2 thinking is encoded in Type 1 memory systems and available to Type 1 thinking*

- Why can't Type 1 processes take care of everything?

1. FEEDBACK

- may enable the integration of information from different mental subsystems
  - subsystems can then share information by generating speech or sensory imagery, thereby making it available to perceptual systems and to the rest of the mind (Dennett 1991)
  - natural language is best for being a content integrator, since most mental subsystems have access to the language system (Carruthers 2006)

2. IMAGISTIC FEEDBACK

  - not random but intentionally controlled, directed to solving some specific problem and guided by learned procedures and tricks
  - we learn ways of constructing verbal arguments and exploiting sensory imagery
  - learning involves myriad micro-changes to the biological brain, encoding new beliefs, skills, habits

3. EXPLOITING EXISTING KNOWLEDGE IN NEW WAYS

  - memories encode a vast amount of information, all potentially relevant to any problem we face
  - autostimulation has a strong selectional effect: asking ourselves a question, many different items of knowledge compete for articulation
    → winning ones prime the next round of selection …

**BY AUTOSTIMULATING, WE CAN HACK A PATH THROUGH THE INFORMATIONAL JUNGLE,**
**MAKING NEW CONNECTIONS & ARRIVING AT NEW CONJECTURES**

# QUALITATIVE DISTINCTION BETWEEN TWO TYPES OF THINKING

**INTUITIVE**

autonomous processes guiding everyday behavior in familiar environments

**REFLECTIVE**

intentional reasoning needed to deal with novel or complex problems

- creating overt representations, questioning ourselves, imagining relevant scenes & objects
- constructing arguments in inner speech
- objects & imagery act as autostimulations
  - providing fresh inputs to our autonomous processes
    (which may then generate a response in the form of more inner speech, other sensory imagery, or an emotional reaction)

SOLUTION

- engaging in cycles of autostimulation & response → we can work our way through problems that would otherwise be beyond us
- culturally transmitted habits of autostimulation create a new level of mental activity, a virtual mind, which engages in reflective thinking

**BY THIS VIRTUAL SYSTEM IN OUR HEADS, WE COME TO APPROXIMATE TO GENERAL INTELLIGENCE**

# ENHANCING HUMAN INTELLIGENCE

## IMPLICATIONS FOR THE PROJECT OF ARTIFICIALLY ENHANCING HUMAN INTELLIGENCE

Which system are we thinking of enhancing? The biological or the virtual mind?

### Enhancing the biological mind

directly interfering with the hardware of the brain

- boosting cognitive functioning with nootropic drugs, neurostimulation, or genetic manipulation
- extending perceptual capacities by hooking up artificial sensors to sensory cortices (relying on brain's plasticity)
- creating artificial cognitive subsystems & implementing such self-organizing systems

None of these technologies will be easy to develop, and installing them will require a detailed understanding of brain functioning and development.

### Enhancing the virtual mind

virtual mind is already a cognitive enhancement

- a set of tricks for extending the powers of the biological brain
- often through the use of artefacts
- tempting to link their emergence with the 'cultural explosion' 30–60,000 years ago, when art, religion, and complex technology first appeared (Mithen 1996)

We might say that the first technological singularity occurred in the Upper Palaeolithic.

### the virtual mind can easily be enhanced

software level: human education = kind of enhancement | adding new hardware: a fundamentally autostimulatory process

**autostimulatory routines we run in our heads are on a par with public ones*

*(intentionally produce & manipulate artefacts & symbols in order to transform complex problems into simpler ones that our biological minds can solve)*

**Technology can vastly extend this process**

**by transforming difficult abstract problems into easy practical ones**

SOLUTION OF THE WHOLE PROBLEM:=

cycles of internal Type 1 processing

+

external electronic processing (= temporally & spatially extended Type 2 processes)

# EXAMPLES

*using a calculator to solve a complicated mathematical problem*

- calculator provides us with new stimuli, creating new, simpler subproblems
- solutions to simpler problems are provided by our Type 1 processes

*supplementing our biological memories with external sources of knowledge*

- tables, reference books, databases

→ consulting an external resource

→ retrieving items of information for use in Type 2 reasoning

- no significant difference between biological memory & external information sources
- both are resources we intentionally access

External resources merely expand the hardware on which the virtual mind is run.

such enhancements to the virtual mind

- easy to install
- interface naturally with our biological minds through our hands & sense organs

→ *just plugging in new cognitive aids via sensory interfaces*

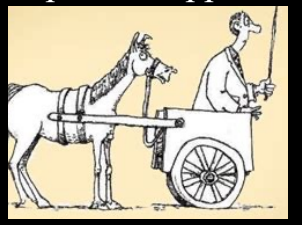
- some training needed in using the devices & interpreting their outputs

For thousands of years, we humans have been enhancing our Type 2 thinking with artefacts, from writing instruments and abacuses through to iPhones and smart glasses, and this sort of enhancement looks set to progress rapidly in coming decades (Clowes 2017, 2019).

# ARTIFICIAL INTELLIGENCE

## DIFFERENT TRADITIONS IN AI CAN BE SEEN AS FOCUSING ON DIFFERENT MENTAL SYSTEMS

a top-down approach



trying to create an operating system without having the hardware

computational modelling of general intelligence focusing on System 2

- proved notoriously intractable: a virtual mind may be easy to enhance, it is difficult to create
- in order to reproduce System 2's powers, we would need to reproduce the powers of the biological mind (=engine of System 2)

*general intelligence could be modeled directly from the top down if we find ways of*

- representing all kinds of Type 1 knowledge in a format that allows for their integration in reasoning
- devising procedures for rapidly retrieving contextually relevant items from a vast knowledge base
- creating a powerful general reasoning system performing a wide range of operations
  (*belief fixation & updating, decision-making, planning, causal reasoning, mentalizing, language processing, abductive inference, creative thinking*)

*unclear what the target of the project would be*

- If we focus on our own idiosyncratic, species-specific, and culture-specific forms, we create artificial versions of ourselves.
  - Human Type 2 thinking
    = shaped by nature & capacities of the specialist subsystems, by the cultural resources, by individual differences

**embodied, behavior-based approaches on System 1**

bottom-up approach: create independent creatures with Type 1 minds & make them developing Type 2 minds

- equip them with goals, social instincts, suites of perceptual, cognitive, motivational systems, communication system
- tuning their goals in the right way → creatures start cognitively stimulating each other & autostimulating →Type 2 thought

*BUT it is unlikely, that we could ensure this outcome through engineering alone.*

- creatures would need to develop social institutions & cultural practices required for Type 2 thinking
- we need to act as guides and teachers (sharing our mental software)

# MAYBE

**The best way to create general intelligence will be to create beings who can create it for themselves.**

If so, then AIs will also have two minds, though the shape of both will probably be quite different from ours.

The form of Type 2 thinking is determined by the nature of the autostimulatory mechanisms employed (the language system, perceptual and imagistic abilities, working memory capacity, and so on), and the virtual minds of AIs might be much richer and more complex than ours.

# The Risks of Enhancement and AI

## LOSING CONTROL or BECOMING DEPENDENT

common alarmist worry

- having embarked on the creation of artificial intelligence, we may lose control of the process

→ our creations may take control of their own development, pursue their own projects, and become indifferent or hostile to us

BUT

This is unlikely to be a fast or straightforward process.

a more feasible strategy

- create artificial creatures with animal-like intelligence
- help them to bootstrap themselves into general intelligence though cultural processes
- nurture them laboriously through a long childhood, both as an artificial species and as individuals

much more pressing concern

- it is easy to enhance our virtual minds, using artefacts to transform problems & supplement our biological memories
- our modern minds are heavily dependent on external support
  - by offloading cognitive drudgery onto electronics in the way that previous generations offloaded manual labor onto mechanical appliances

| LABORIOUS BUSINESS | SHORTCUT |
|---|---|
| programming our biological brains to support Type 2 thinking | technology is offering increasingly powerful new cognitive aids |
| learning to do long division | learn to use a calculator |
| memorizing of historical facts | learn to access online encyclopedia |
| memorizing spellings | learn to run a spellcheck |

# The Risks of Enhancement and AI

**expecting technology to give us many completely new capacities**

- supplementing our biological minds with external modules
- using them for information, entertainment, motivational stimuli, to make visual, aural, tactile contact with far-off people & places
- radically enriched conscious minds
- allowing us to develop new ways of working, socializing, loving

| obvious advantages & impossible to resist | obvious dangers & if they fail, it won't be easy to fall back on older technology |
|---|---|
| Why should a lawyer spend years studying case law if they can buy a tiny earpiece that will instantly retrieve contextually relevant data as needed and feed it to them? | conscious minds & biological brains become dependent on external electronic hardware<br>• brains are robust, well-protected organs (product of millions of years of natural R&D & have a remarkable capacity for self-repair)<br>• electronic devices are far more vulnerable (a solar flare might knock them out and leave us cognitively disabled) |

**being at the mercy of those who control the technology**

having offloaded so much of our skill and knowledge

→ no resources to assess the value of the information & guidance we are fed

→ easy to be manipulated

already there: use of social media bots to manipulate opinion

- seemingly relevant images & bits of information pop up on social media, guiding our thoughts & decisions

**Don't worry about master AIs but worry about the human servants**

avoid developing our virtual minds to the point where they are no longer really ours | no longer tethered to our biological minds & purposes & values

**PARADOX OF THE VIRTUAL MIND**

In learning how to manipulate our biological minds and create virtual minds for ourselves, we risk undermining the locus of purpose and control that our biological minds sustained. It is the price of being creatures with two minds.

# References


- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*(1–3), 139–159. https://doi.org/10.1016/0004-3702(91)90053-M.
- Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Oxford University Press.
- Carruthers, P. (2009). An architecture for dual reasoning. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 109–127). Oxford: Oxford University Press.
- Carruthers, P. (2015). *The centered mind: What the science of working memory shows us about the nature of human thought*. Oxford: Oxford University Press.
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.
- Clowes, R. W. (2017). Extended memory. In S. Bernecker & K. Michaelian (Eds.), *The Routledge handbook of philosophy of memory* (pp. 243–254). Abingdon: Routledge.
- Clowes, R. W. (2019). Immaterial engagement: Human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences, 18*(1), 259–279. https://doi.org/10.1007/ s11097-018-9560-4.
- De Neys, W. (Ed.). (2018). *Dual process theory 2.0*. New York: Routledge.
- Dennett, D. C. (1991). *Consciousness explained*. New York: Little, Brown.
- Dennett, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. New York:W.W. Norton.
- Descartes, R. (1984). *The philosophical writings of Descartes: Volume 2* (Trans., J. Cottingham,
- R. Stoothoff, & D. Murdoch). Cambridge: Cambridge University Press.
- Diaz, R. M., & Berk, L. E. (Eds.). (1992). *Private speech: From social interaction to self- regulation*. Hillsdale: Lawrence Erlbaum.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist, 49*(8), 709–724.
- Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hove: Lawrence Elrbaum Associates.
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378–395.

# References


- Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.
- Evans, J. S. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press. Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241. https://doi. org/10.1177/1745691612460685.
- Frankish, K. (1998). Natural language and virtual belief. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 248–269). Cambridge: Cambridge University Press.
- Frankish, K. (2004). *Mind and supermind*. Cambridge: Cambridge University Press.
- Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal dis- tinction. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford: Oxford University Press.
- Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass, 5*(10), 914–926. https://doi.org/10.1111/j.1747-9991.2010.00330.x.
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies, 23*(11–12), 11–39.
- Frankish, K. (2018). Inner speech and outer thought. In P. Langland-Hassan & A. Vicente (Eds.), *Inner speech: New voices* (pp. 221–243). Oxford: Oxford University Press.
- Frankish, K., & Evans, J. S. B. T. (2009). The duality of mind: An historical perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1–29). Oxford: Oxford University Press.
- Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory & Psychology, 20*(6), 733–743. https://doi.org/10.1177/0959354310378184.
- Gleick, J. (1992). *Genius: The life and science of Richard Feynman*. New York: Pantheon Books. Heyes, C. M. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge, MA: Harvard University Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Keren, G., & Schul, Y. (2009). Two is not always better than one. *Perspectives on Psychological Science, 4*(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on com- mon principles. *Psychological Review, 118*(1), 97–109. https://doi.org/10.1037/a0020762.
- Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences, 22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(02), 57–74.

# References


- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(02), 57–74. https://doi.org/10.1017/S0140525X10000968.
- Mirolli, M., & Parisi, D. (2011). Towards a Vygotskyan cognitive robotics: The role of language as a cognitive tool. *New Ideas in Psychology, 29*(3), 298–311. https://doi.org/10.1016/j.newideapsych.2009.07.001.
- Mithen, S. J. (1996). *The prehistory of the mind: A search for the origins of art, religion and sci- ence*. London: Thames & Hudson.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988–1010.
- Osman, M. (2018). Persistent maladies: The case of two-mind syndrome. *Trends in Cognitive Sciences, 22*(4), 276–277. https://doi.org/10.1016/j.tics.2018.02.005.
- Pennycook, G., Neys, W. D., Evans, J. S. B. T., Stanovich, K. E., & Thompson, V. A. (2018). The mythical dual-process typology. *Trends in Cognitive Sciences, 22*(8), 667–668. https://doi. org/10.1016/j.tics.2018.04.008.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123–205.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah: Lawrence Erlbaum Associates.
- Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
- Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.
- Stanovich, K. E. (2009b). *What intelligence tests miss: The psychology of rational thought*. New Haven: Yale University Press.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of type 1 and type 2 processing. *Mind & Society, 11*(1), 3–13. https://doi.org/10.1007/s11299-011-0093-6.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(05), 645–726.
- Steels, L., & Brooks, R. A. (Eds.). (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Hillsdale: Lawrence Erlbaum Associates.
- Vygotsky, L. (1986). *Thought and language* (Trans. & Ed., A. Kozulin). Cambridge, MA: MIT Press.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Winsler, A., Fernyhough, C., & Montero, I. (Eds.). (2009). *Private speech, executive functioning, and the development of verbal self-regulation*. Cambridge: Cambridge University Press.