NACS 645 – Two systems to decide

-

Valentin Guigon

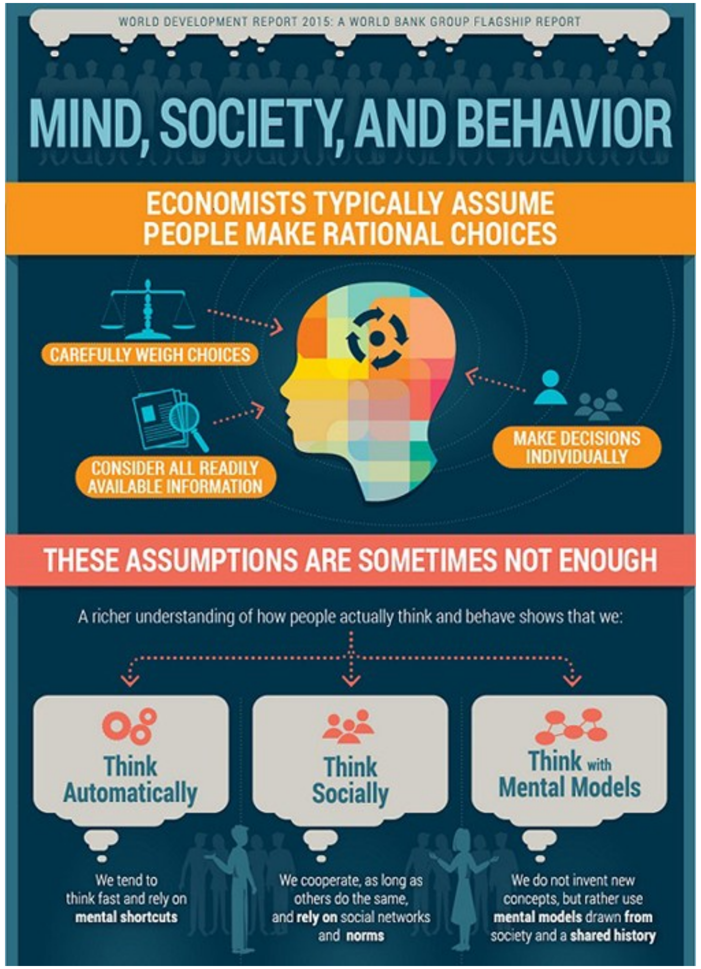
DEPARTMENT OF PSYCHOLOGY

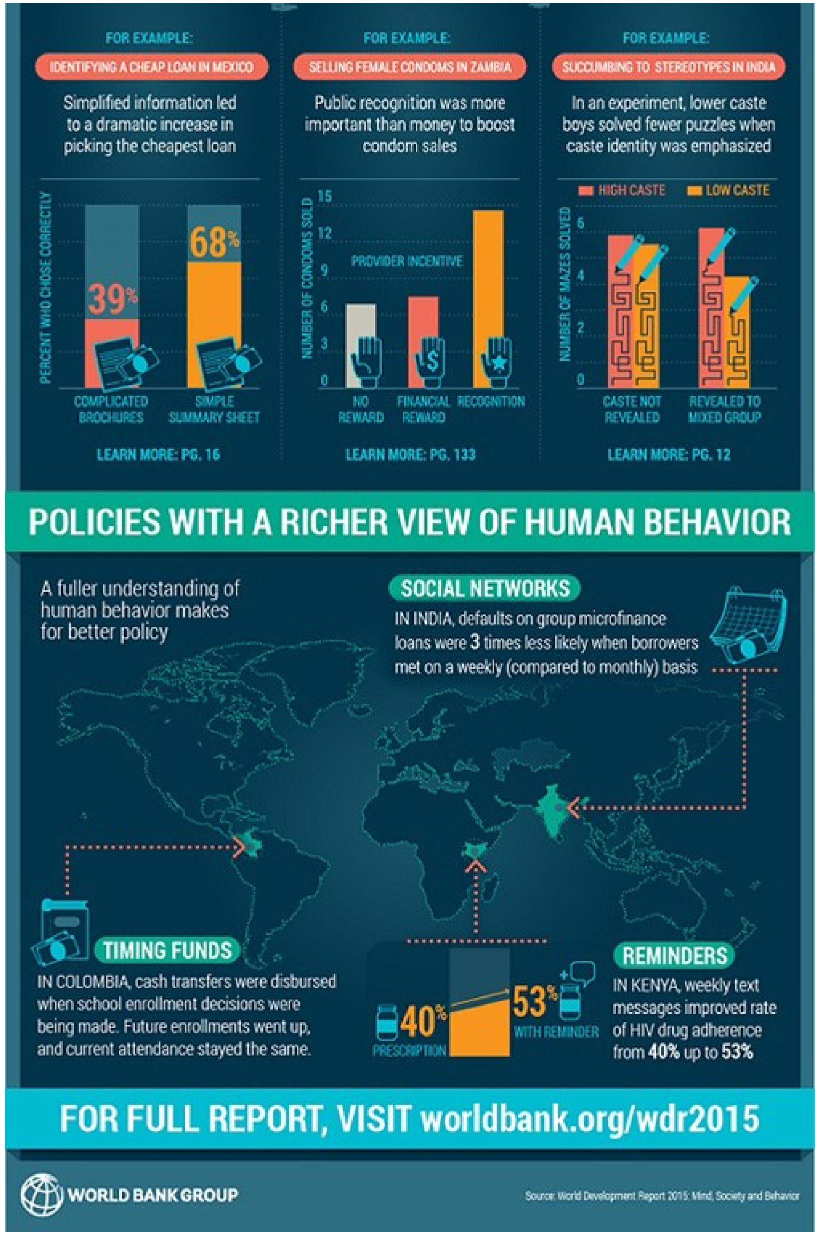
SLD Lab

PROGRAM IN NEUROSCIENCE & COGNITIVE SCIENCE

System 1 and System 2 - A big idea

In 2015, the *World Bank* called on decision-makers to use System 2 thinking in order to avoid the errors associated with System 1 thinking





White; World Bank, *World Development Report 2015: Mind, Society and Behaviour*

Modern traditions of rationality

- Logical rationality (post-WWII, Cold War, Game theory): formal, context-independent norms grounded in **choice/consistency axioms** (e.g., completeness, transitivity, independence) and optimization (expected utility maximization, Bayesian updating, Nash equilibria)
  - Provides the norms
- Heuristics-and-biases program (70s): humans **follow rules of logical rationality**, but they **tend to deviate** from them; these deviations are called biases; biases justify interventions (e.g., nudges)
  - Diagnoses deviations *relative to those norms*
- Ecological rationality (90-2000): **rationality is not consistent with classical axioms** (transitivity, coherence, stability) but rather **bounded** by biological and ecological constraints: uncertainty, task complexity and available cognitive resources
  - Questions the scope of the norms

Kahneman, 2022. « Biographical ». The Nobel Prize.

The two thinking systems

Early work on probabilistic reasoning

Observations that kids and adults have good statistical intuitions (e.g., Piaget & Inhelder; Edwards)

Some critics (Gigerenzer, 2025):

- Repeated trials with random devices (urns, dice)
- Performance assessed across many observations
    - Opportunities for learning, calibration, and error correction

1970s Heuristics-and-Biases

Observations of systematic deviations from logical and probabilistic norms

Some critics (Gigerenzer, 2025):

- Numerical random devices replaced by verbal scenarios
- Single-shot judgments on text
    - No opportunity for learning

Implicit assumptions

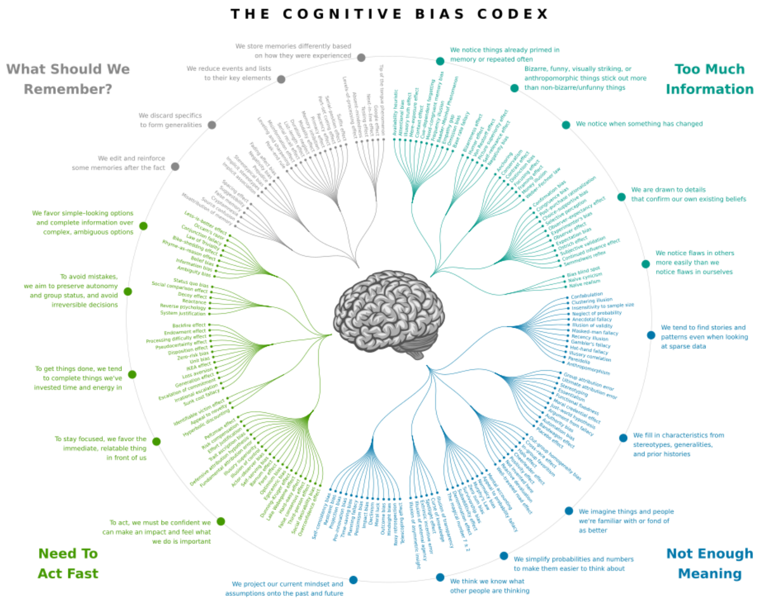
- One correct answer per problem
- Normative standard fixed in advance

System 1 / System 2 as a resolution

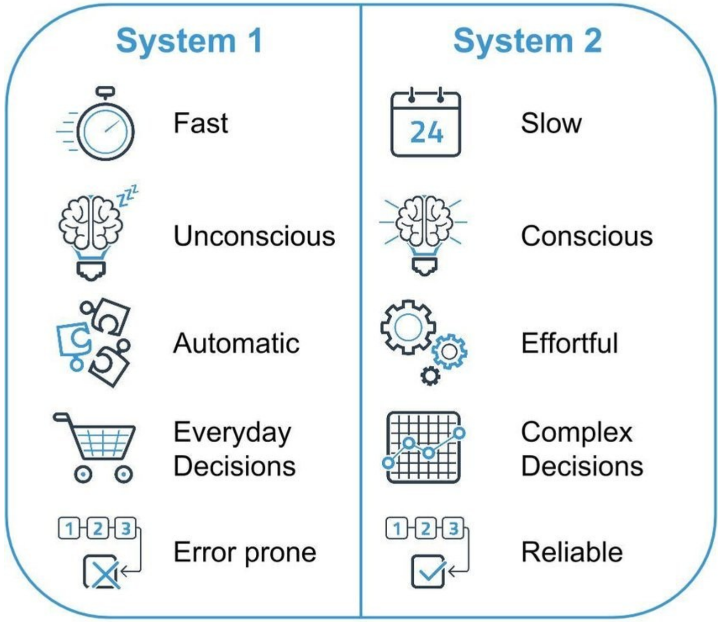
Errors of intuition occur when System 1 generates the error and System 2 fails to correct

- System 1: fast, intuitive, automatic → source of deviations
- System 2: slow, deliberative, controlled → potential corrector
- Based on Sloman; reconciles cognitive biases with the fact that those same biases could be made disappear

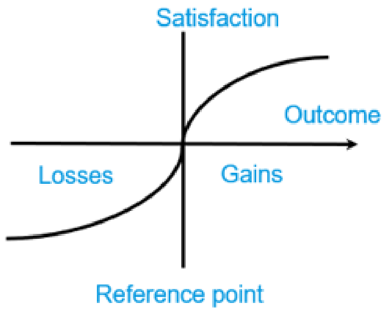
The heritage of Kahneman & Tversky

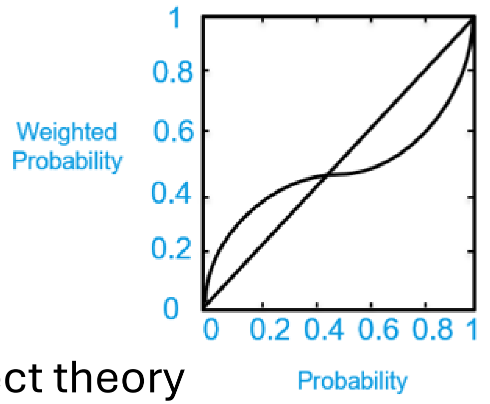


Heuristics and biais
1970's



Dual-process theory
1990-2000





Prospect theory
1979

System 1 and System 2 - Framework

Trait	Type 1 (Fast)	Type 2 (Slow)
Consciousness	Unconscious	Conscious
Intentionality	Unintentional	Intentional
Efficiency	Efficient (low cognitive cost)	Inefficient (high cognitive cost)
Controllability	Uncontrolable	Controlable

Postulates:

- Mental processes naturally fall into **two distinct types**
- **Knowing one trait** (e.g. a process is unconscious) **helps deducing the other traits** (e.g. therefore is also unintentional, efficient and uncontrollable).
- This grouping **reflects the human cognitive architecture**

Melnikoff et Bargh (2018):

- Many psychological processes **combine traits from both types**. e.g., Intentional but unconscious: driving, typing, playing the piano
- The traits themselves are not **unitary**, each breaking down into **sub-components** that don't always coincide. e.g., controllability is heterogeneous (modulable vs preventable)

Dual-process theories

The tradition of dual-process theories of reasoning originates from classical views of rationality & philosophy: ***passion vs reason***, vs *Damasio*

Some other modern dual-process theories:

- Emotions: hot vs cold
- Action-selection: habitual vs goal-directed
- Persuasion: central route vs peripheral route
- Reinforcement learning: model-free vs model-based
- Decision: visceral/somatic markers vs high-level cognition

Problem: Dual-process theories are good at organizing intuitions, but weak at constraining explanations. It’s easy to “maintain” beliefs about theories when they are not properly constrained:

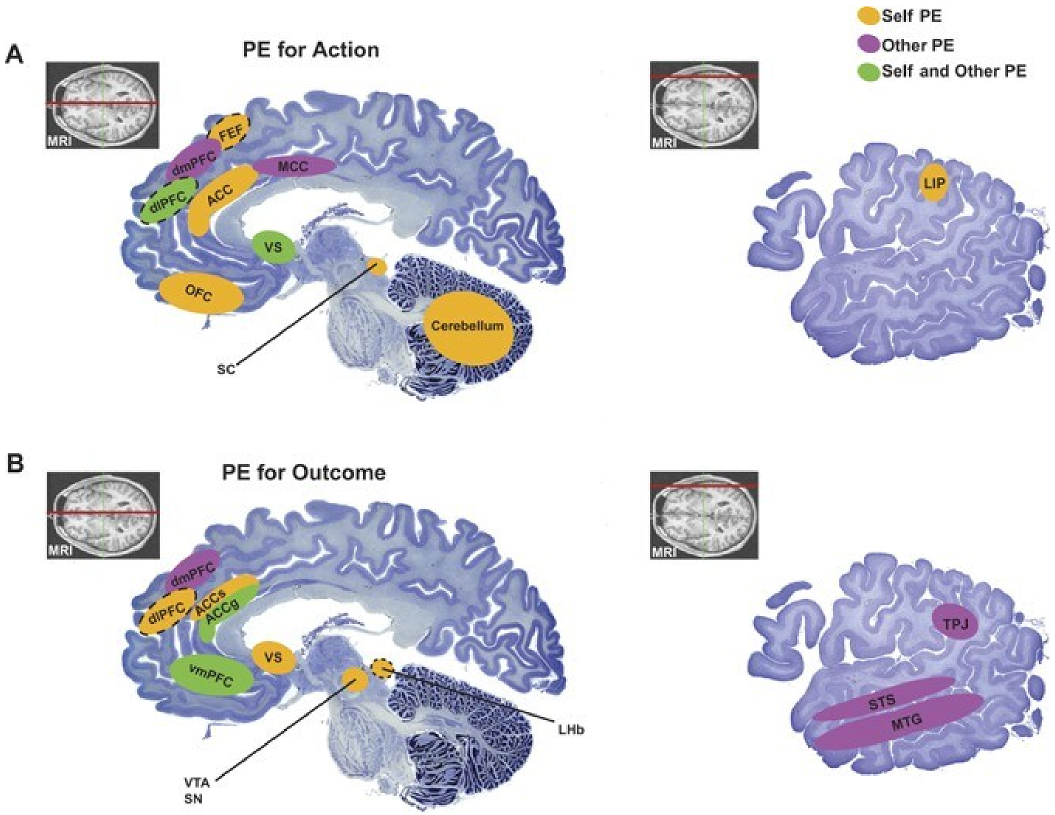
e.g., modularity: Fodor’s 9 rules -> $9 - n$ rules qualitative models;

e.g., system 1/system 2 traits revised

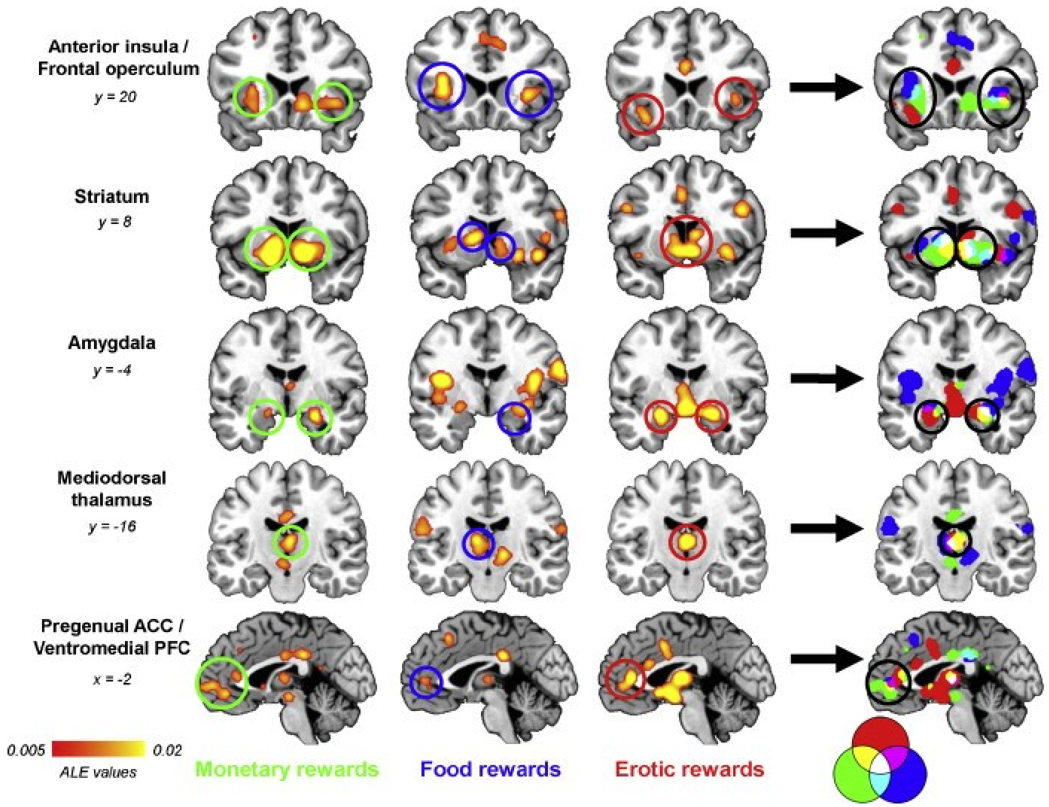
Joiner et al., 2017. *NPJ Science of learning*
Sescousse, Dreher et al., 2013. *Neuroscience & Biobehavioral Reviews*

Emotion vs Reason

Neural overlap undermines **moralized mappings** of dual-process theories



Social prediction



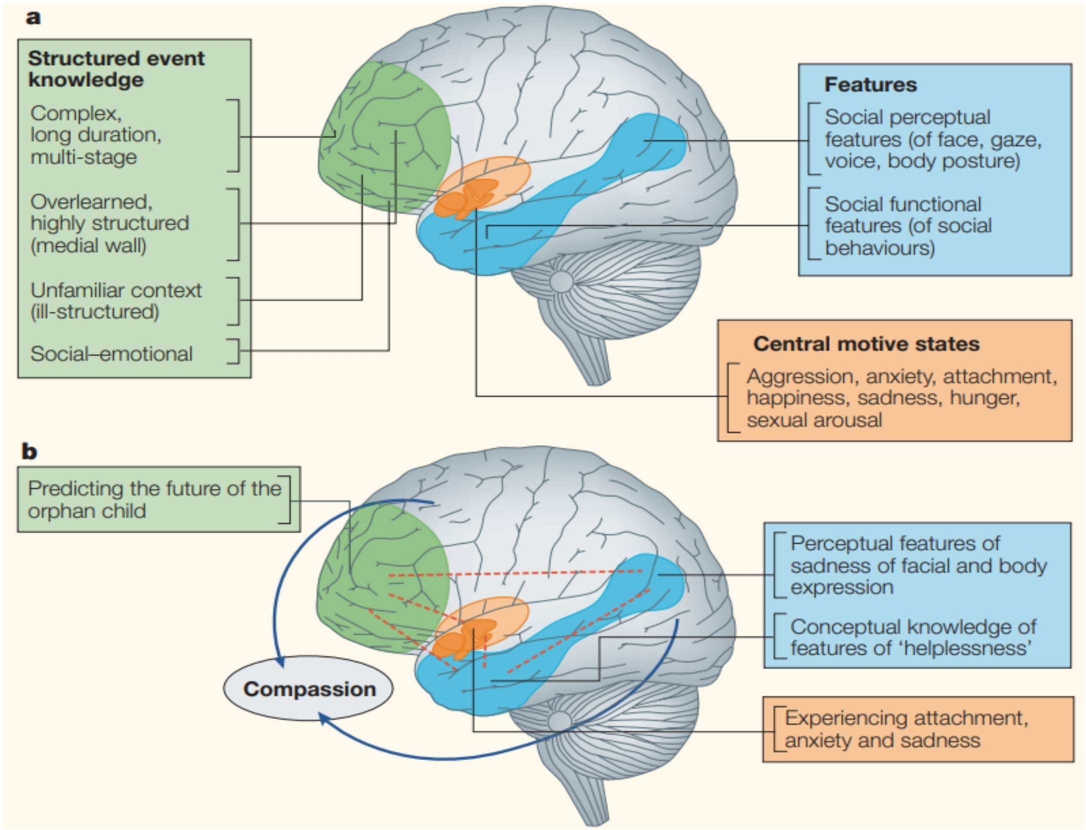
Common reward circuit during outcome

- Processes often labeled as “emotional” participate in prediction, learning, and valuation
- These are core computational functions, not sources of error per se
- **The brain does not implement a clean separation between fast/affective vs slow/rational systems**

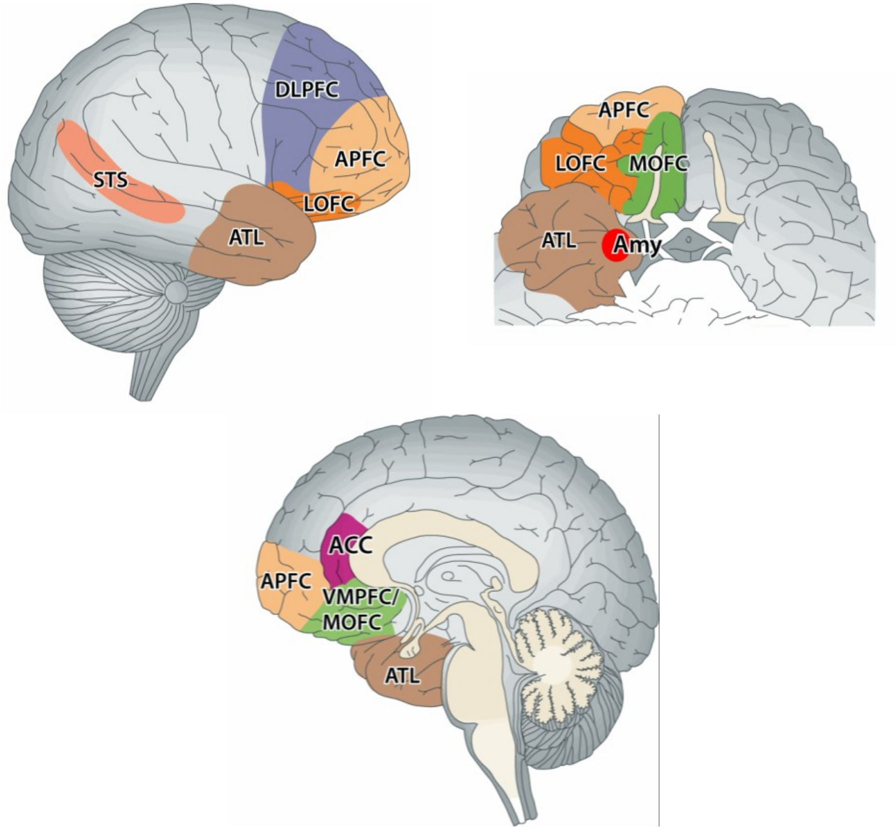
Moll et al., 2005. *Nature reviews neuroscience*
Forbes et Grafman, 2010. *Annual review of neuroscience*

Automatic vs controlled

In the case of morality, automatic and slow processing both rely on overlapping complex and modular structures



Moral processing



Moral judgments

Even in domains often used to illustrate dual-process conflicts (morality, emotion, control), **the brain does not respect the clean fast/slow or emotion/reason divide**

The Good and the Bad

Griffits et Tenenbaum, 2006. *Psychological Science*
Callaway et al., 2018. *Proceed. of the Ann. Meet. of the Cog. Sci. Society*
Melnikoff et Bargh, 2018. *Trends in Cognitive Sciences*

Some dual-process theories don't make sense anymore (reason vs emotions).
What about dual-process theories of reasoning?

Judgments typically associated with System 1 can be optimal:

- Everyday decisions approach the performance of an ideal Bayesian observer (Griffiths & Tenenbaum, 2006)
- Regular planning of tasks achieve ~86% of the optimal trade-off between decision quality and cognitive cost (Callaway et al., 2020)

Judgments typically associated with System 2 can lead to motivated beliefs:

> *Calling Type 2 thinking good is to champion motivated reasoning, the domain of self-serving rationalizations and of finding creative self-serving justifications [...]. (Melnikoff et Bargh, 2018)*

Rationality of processes seem to rather depend on task structure, uncertainty, goals, and computational constraints:
The quality of a judgment doesn't strictly map whether it is fast or slow, automatic or deliberate

NACS 645 – Model-free vs Model-based

-

Valentin Guigon

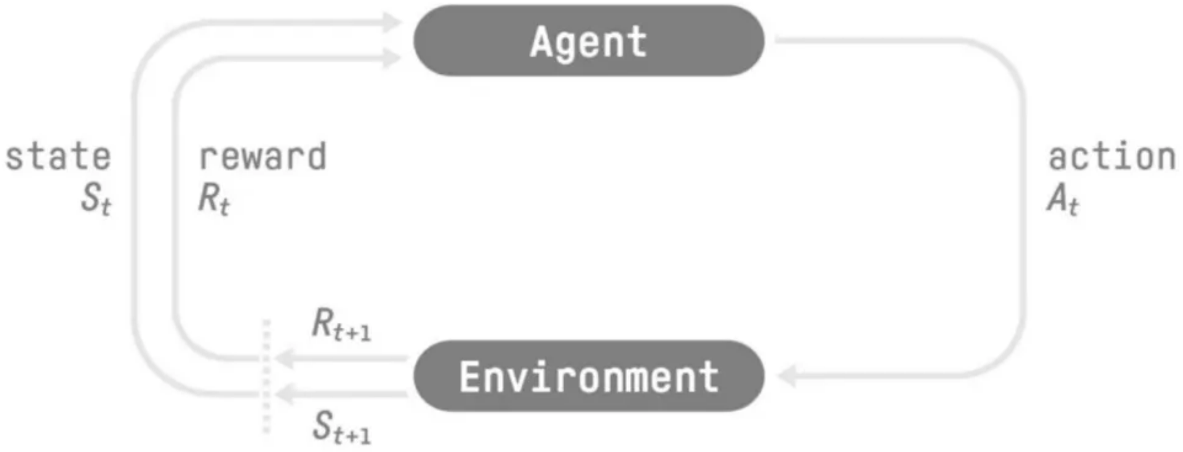
DEPARTMENT OF PSYCHOLOGY

SLD Lab

PROGRAM IN NEUROSCIENCE & COGNITIVE SCIENCE

RL from the algorithms perspective

Learning



All adaptation arises solely from the sequence of states, actions, and rewards generated by interaction with the environment.

Reinforcement learning is a framework for solving decision problems.
Agents learn from the environment by interacting with it through trial and error.
They receive rewards/punishments as feedback (no supervision).

Repeatedly:

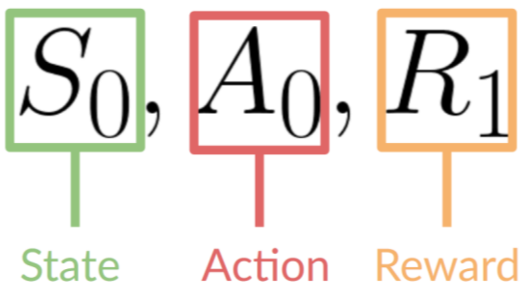
- The agent receives **state** S_0 from the **Environment**
- Based on that **state** S_0 **,**the Agent takes **action** A_0
- The environment goes to a **new state** S_1
- The environment gives some **reward** R_1 to the Agent

Through this trial-and-error cycle, the agent adjusts its behaviour to **maximize expected cumulative reward** (reward hypothesis)**.**

Markov Decision processes

A Markov Decision Process (MDP) is the standard formalism for reinforcement learning.

It defines an environment in terms of:



The **Markov property** states that optimal decisions depend only on the *current* state, not on the full history.
The agent's objective is to select actions that maximize the **discounted cumulative reward**, weighting future rewards less than immediate ones.

- **Observation/States:** information describing the current situation
    - *Fully observed* setting: a **state** provides a **complete description of the world** (chess)
    - *Partially observed* setting: an **observation** provides a **partial description of the state** (poker)
- **Actions:** the set of all moves the agent can take
    - Discrete: the number of possible actions is **finite**
    - Continuous: the number of possible actions is **infinite**
- **Rewards:** scalar indicating the immediate consequence of an action

Huggingface, Deep RL Course.

Rewards and discounting

Cumulative reward at each time step t

$$\underbrace{R(\boxed{\tau})}_{\text{Return: cumulative reward}} = r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4} + \dots$$

Return: cumulative reward

Trajectory (read Tau)
Sequence of states and actions

$$R(\tau) = \sum_{k=0}^{\infty} r_{t+k+1}$$

Discounted expected cumulative reward
at each time step t

$$R(\boxed{\tau}) = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \dots$$

Return: cumulative reward

Gamma: discount rate

Trajectory (read Tau)
Sequence of states and actions

$$R(\tau) = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

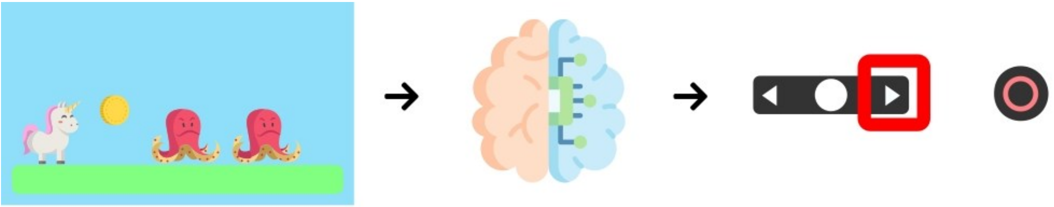
If the agent only exploits that they learnt so far, they may neglect alternatives with higher payoff, and not maximize their expected reward.
Hence, there is an **exploration/exploitation tradeoff** the agent should solve.

Huggingface, Deep RL Course.

How to maximize expected cumulative reward?

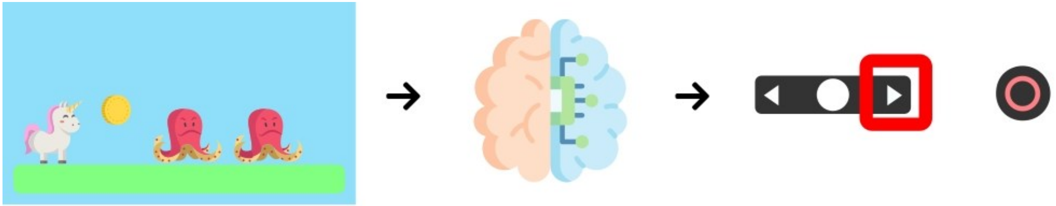
Reinforcement learning aims to find an optimal policy π^* (a mapping from states to actions).
There are **two ways** to reach this policy: policy-based methods and value-based methods

Policy-based: deterministic

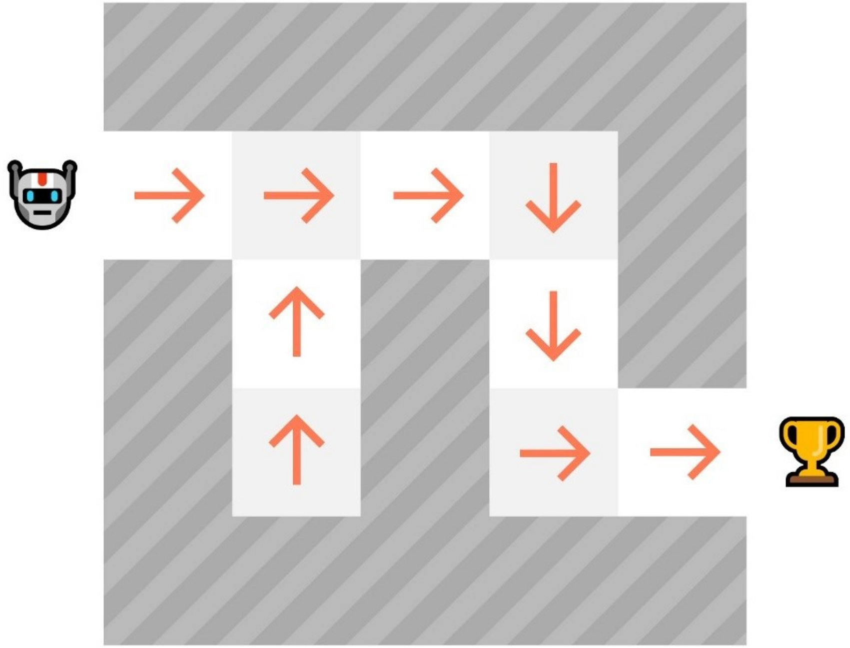


State s_0 → $\pi(s_0)$ → a_0 = Right

Policy-based: stochastic (allows exploration)



State s_0 → $\pi(A|s_0)$ → a_0 = [Left: 0.3, Right: 0.7]



Value-based methods

Policy-Based Methods
- Learn the **policy directly** (*state -> action* mapping)
- The model outputs the **action** (or a probability distribution over actions) **for each state**
- The agent improves this mapping through experience

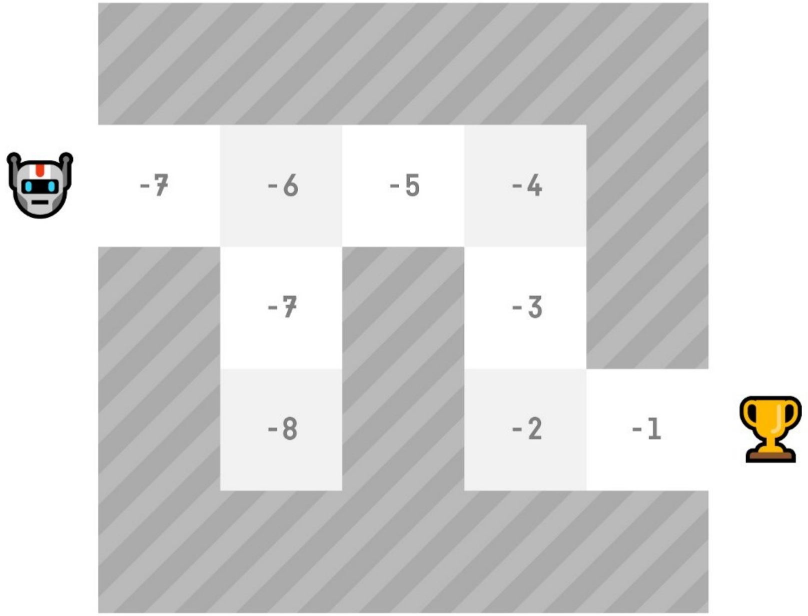
The agent "learns how to act" without explicitly evaluating states.

Value-Based Methods
- Learn a **value function** (*state -> expected return* mapping)
- V(s) or Q(s,a) estimates "how good" it is to be in a state or to take an action
- The policy is **derived** by choosing actions that lead to higher estimated values

The agent "learns which states are better" and acts by following a policy (i.e., moving toward higher-value states).

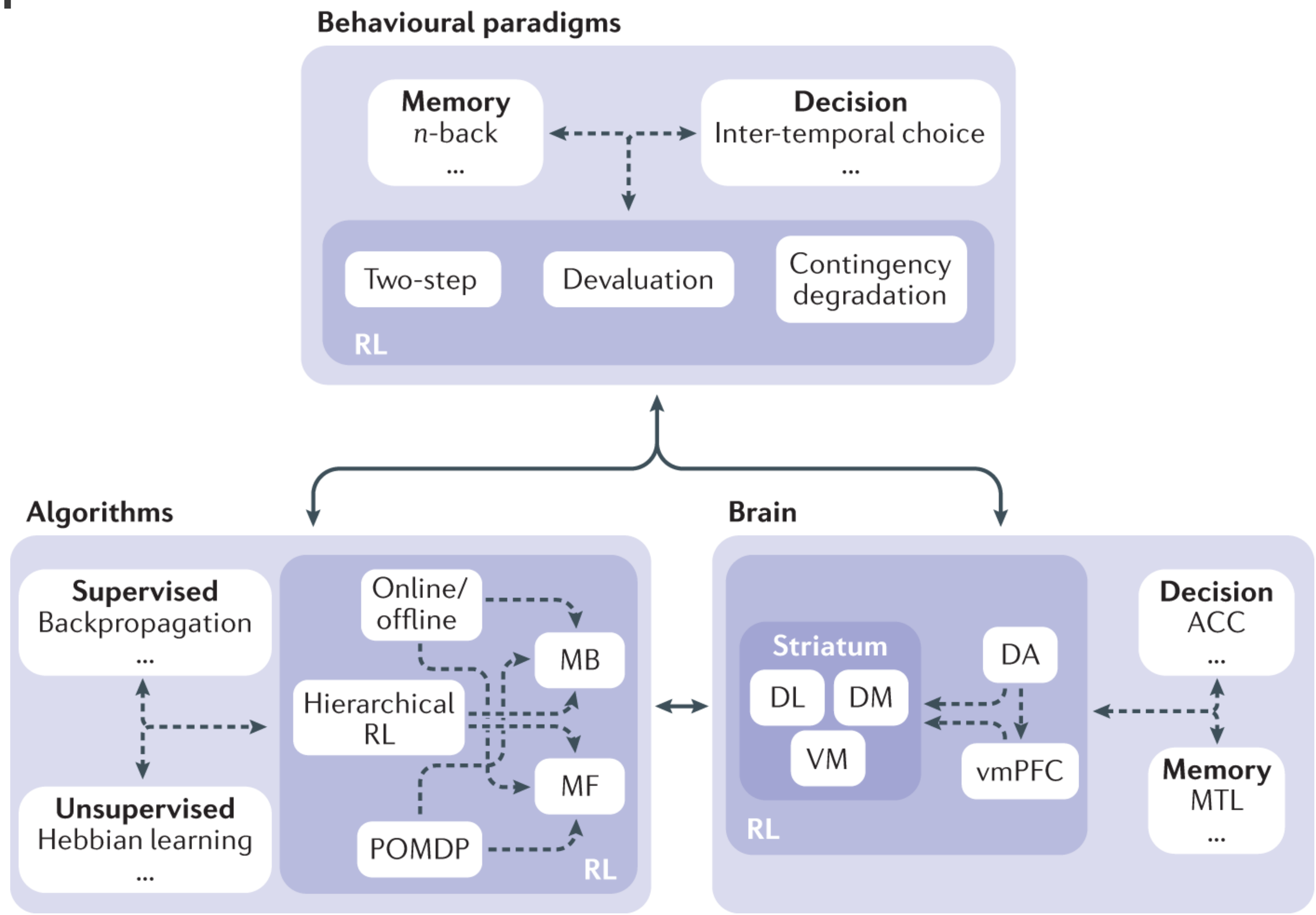
Value-based

$$\underbrace{v_\pi(s)}_{\text{Value function}} = \underbrace{\mathbb{E}_\pi\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots \right.}_{\text{Expected discounted return}} \left| \underbrace{S_t = s}_{\text{Starting at state s}}\right]$$



Collins & Cockburn, 2020. *Nature Reviews Neuroscience.*

RL across fields of research



RL is a formal language for learning-based, sequential decision-making. It gives tools to talk about goals, prediction, uncertainty, credit assignment.

Early behavior studies (1890s–1950s): Trial-and-error learning, conditioning, habit vs goal-directed actions

Algorithmic RL (1950s–1990s): Formal mathematical tools for sequential decision-making: Dynamic programming, MDPs, TD learning, Q-learning

Neural interpretations (1990s–2000s): Dopamine prediction-error signals; striatal habit learning; prefrontal planning. Computational RL imported into neuroscience

Glossary

- **Reinforcement Learning (RL)**
  Framework in which an agent learns through interaction with an environment to maximize expected cumulative reward.
- **Markov Decision Process (MDP)**
  Formal model defining states, actions, transition dynamics, and rewards under the Markov property (current state contains all decision-relevant information).
- **State**
  Complete description of the environment at a time point. In a fully observed setting, no hidden variables remain.
- **Observation**
  Partial information available to the agent when the environment is not fully observed.
- **Action**
  A choice available to the agent. Action spaces may be discrete or continuous.
- **Transition Function**
  Specification of how states evolve when actions are taken: $P(s' \mid s, a)$.
- **Reward Function**
  Scalar feedback indicating the immediate consequence of an action in a state.
- **Discounted Return**
  Sum of future rewards, discounted to give more weight to near-term outcomes.
- **Policy (π)**
  A decision rule mapping states to actions, deterministic or stochastic.
- **Policy-Based RL**
  Learn the policy directly (state → action). Optimization focuses on improving behaviour.
- **Value-Based RL**
  Learn a value function (state → expected return, or state–action → expected return). The policy is derived from these values.
- **Value Function (V, Q)**
  $V(s)$: expected return from state s.
  $Q(s,a)$: expected return from taking action a in state s and following the policy.
- **Model-Free RL (MF)**
  Learns value estimates or policies directly from experience without an explicit transition model.
- **Model-Based RL (MB)**
  Learns or uses a model of transitions and rewards to evaluate future outcomes via planning.
- **Markov Property**
  The current state contains all the information needed for optimal decision-making; the past is irrelevant given the present.

NACS 645 – Collective knowledge

-

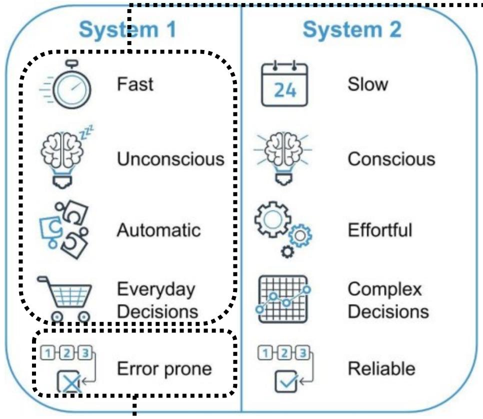
Valentin Guigon

DEPARTMENT OF PSYCHOLOGY

SLD Lab

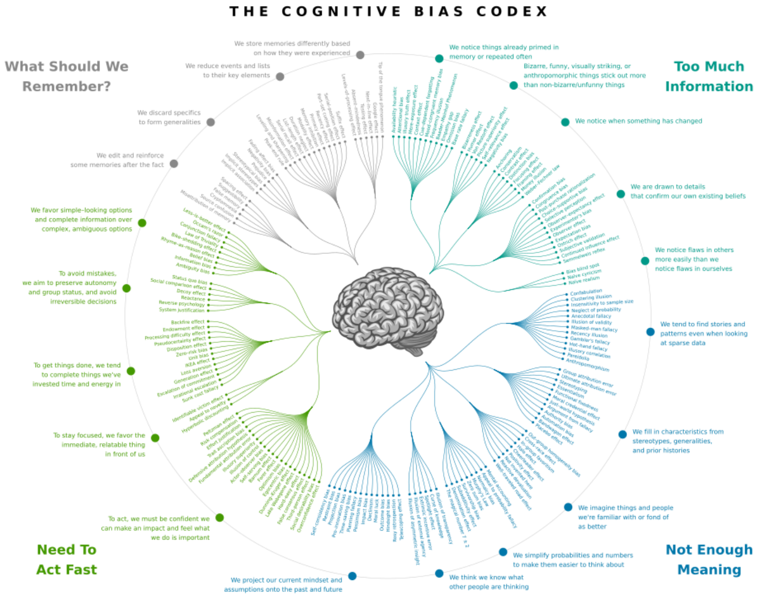
PROGRAM IN NEUROSCIENCE & COGNITIVE SCIENCE

System 1 & System 2
Heuristics and biases



Heuristics

- **Simple, fast, frugal** cognitive strategies
- Often ignore part of the information to find a good-enough (rather than perfect) solution



Biases

- Often seen as **error or systematic deviations** between a human judgment and a norm of rationality (e.g., laws of probabilities or logic)

The role of biases in judgments

When information is scarce, degraded, uncertain, complex, noisy
When the environment is sufficiently predictable

- **Predictive superiority**

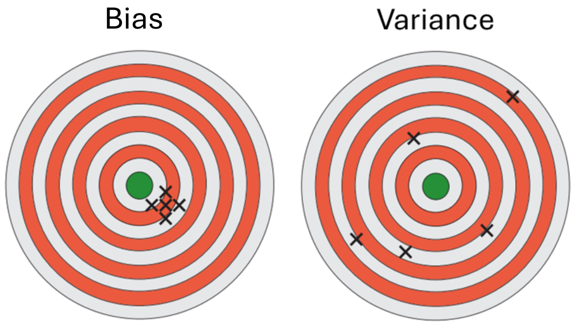
- **Robustness to uncertainty**
  Ignoring information (i.e., selectively using, weighting, or interpreting information) can make predictions less sensitive to noise and small samples

- **Cognitive efficiency**
  Reduces cost while maintaining sufficient performance (Martignon et al., 2008)

- **By simplifying**, heuristics introduce a bias:

→ This bias reduces the instability of predictions (variance)
→ Improves robustness and generalizability to similar situations, especially under uncertainty

Prediction error in predictive models:



+ noise (ε)

Gigerenzer & Brighton, 2009

The role of noise in judgments

Bias

- **Systematic directional shift in judgments** compared to the ground truth

Variance

- Instability of predictions across new situations

Noise

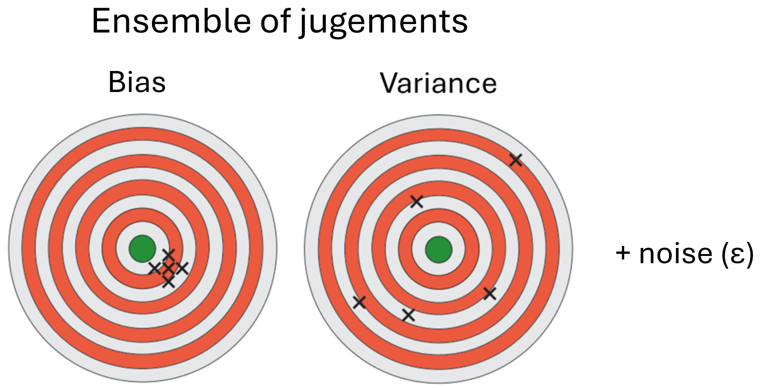
- **Random and unpredictable dispersion** of judgments around the truth
  (n.b., includes patterns observed ex-post,
  excludes rules predicted ex-ante)

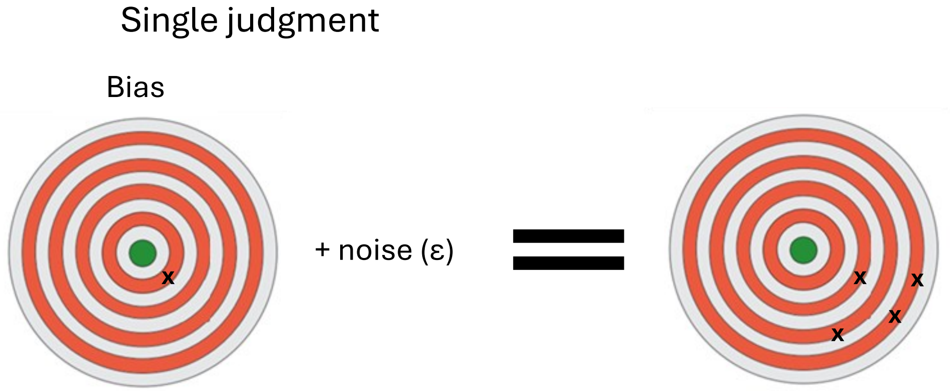
Bias

- **Systematic average deviation …**

Noise

- **Random and unpredictable dispersion …**





Types of noise

- System noise : **overall variability** within a multi-judge system
    - Level noise : **systematic differences between judges** (e.g., some stricter, others more lenient)
    - Pattern noise : **a judge's characteristic fluctuation across specific cases** (i.e., judge x case interaction, aka personal signatures observed post-hoc on particular cases [correlations between data points], but not generated by latent rule)
- Occasion noise : **intra-individual fluctuations**, i.e., within a single judge system (e.g., mood, fatigue, cog depletion, time of day)

Level noise

- E.g., A judge who hands down less severe sentences than peers (Clancy, Bartolomeo et al., 1981)

Pattern noise

- E.g., A judge who is stricter for some crimes and more lenient for others (Clancy, Bartolomeo et al., 1981)

Occasion noise

- Agents more likely to approve bank loans in the morning (decision fatigue reduces evaluation capacity) (Baer and Schnall, 2021)
- Physicians more likely to prescribe opioids at the end of workday (Philpot et al., 2018)
- Internal fluctuations affecting information processing

Kahneman, Sibony, Sunstein, 2021

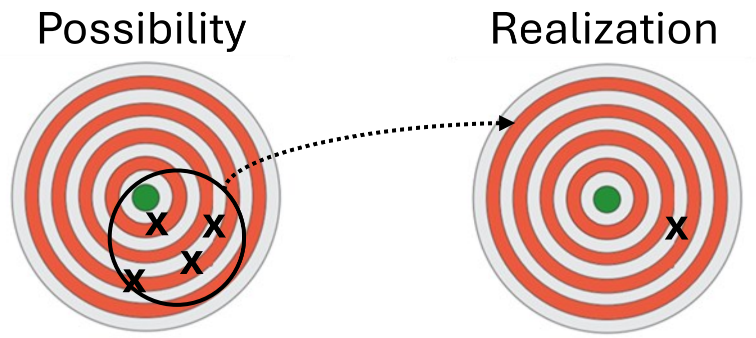
The (wrong) tendency to treat each judgment as isolated

Noise is largely invisible because we handle each judgment as if it were unique.

Noise is a property that emerges only when examining a set of cases.

It is a property of a general process realized during a particular event.

"A singular decision is a recurrent decision that happens only once.» (Kahneman, Sibony et Sunstein)



Kahneman, Sibony, Sunstein, 2021

Noise factors

Individual factors

- Mood, fatigue, personal context
- Experience, preferences, personal beliefs
- Information processing

Cognitive factors

- Overconfidence in one's own judgment or in colleagues' judgments
- Lack of awareness of the problem: noise is rarely visible in everyday practice
- Information overload

Social factors

- Influence from others, group effects, polarization
- Amplification within groups: informational cascades, social influence
- Polarization: reinforcement of opinions through peer interaction

Organizational factors

- Absence of standardized procedures
- Opaque decision processes (e.g., a second evaluator influenced by the first)
- Lack of immediate and clear feedback (medicine, hiring, justice, etc.)
- Preference for harmony and avoidance of disagreement (consensus over honest evaluation)

Kahneman, Sibony, Sunstein, 2021

Moods induce (unpredictable) noise
Emotions induce (predictable) biases

Table 1 Two illustrations of the appraisal-tendency framework, originally developed by Lerner & Keltner (2000, 2001) and updated here.[a] Table adapted from Lerner JS, Keltner D. 2000. Beyond valence: toward a model of emotion-specific influences on judgment and choice. *Cogn. Emot.* 14(4):479, table 1, with permission from the publisher

Cognitive appraisal dimensions	Illustrations: negative emotions		Illustrations: positive emotions	
	Anger	Fear	Pride	Surprise
Certainty	High	Low	High	Low
Pleasantness	Low	Low	High	High
Attentional activity	Medium	Medium	High	Medium
Anticipated effort	High	High	Low	Low
Individual control	High	Low	High	Medium
Others' responsibility	High	Medium	Low	High
Appraisal tendency	Perceive negative events as predictable, under human control, and brought about by others	Perceive negative events as unpredictable and under situational control	Perceive positive events as brought about by self	Perceive positive events as unpredictable and brought about by others
Influence on relevant outcome	**Influence on risk perception**		**Influence on attribution**	
	Perceive low risk	Perceive high risk	Perceive self as responsible	Perceive others as responsible

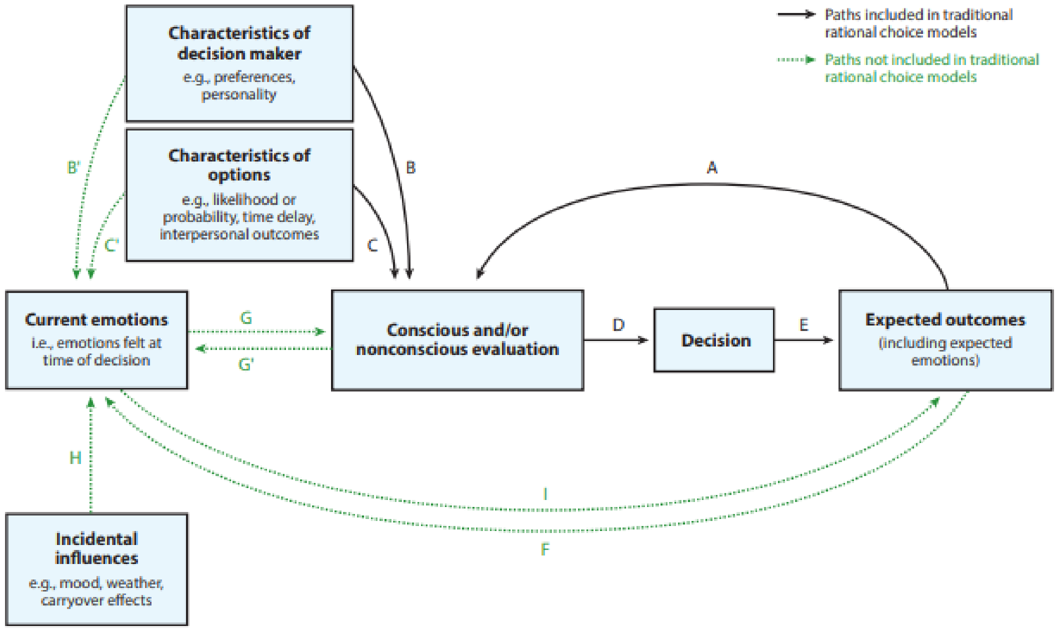


Figure 2
Toward a general model of affective influences on decision making: the emotion-imbued choice model.

How to improve judgments I:
Some principles for decreasing the noise

The purpose of judgment is accuracy, not expression

- Compare the current case with similar cases – treat the case as neither unique nor routine
- Constrain or replace human judgment with simple rules or statistical models (algorithms)
- Calibrate confidence

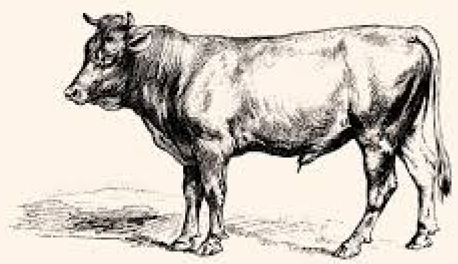
- Have judgments made independently and privately
- Aggregate independent judgments – weight them (mean or other methods) to smooth individual variability
- Weight by expertise
- …

Kahneman, Sibony, Sunstein, 2021

How to improve judgments II: Wisdom of the crowd

1. Wisdom vs Stupidity of crowds

- Aggregated judgment from a large group is more accurate than that of a single expert
- Requires private, independent, and diverse predictions



- 787 estimations
- Mean: 1197 lbs
- Median: 1207 lbs
- Real weight: 1198 lbs

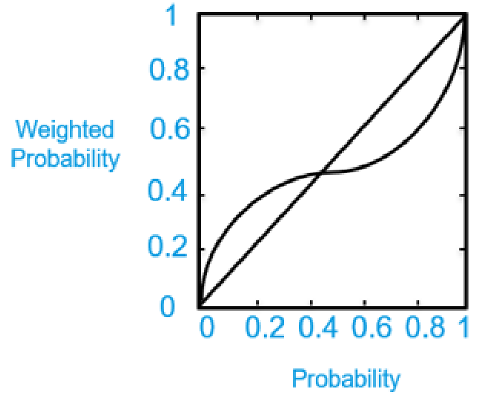
If participants copy one another or anchor their beliefs on shared erroneous information,
the market can converge toward incorrect values

How to improve judgments III: Incentivizing

2. Belief elicitation through monetary incentives:

(skin in the game) (Kant; Schotter et Trevino, 2014):

- Promotes accurate estimates and calibrated confidence



3. Efficient markets:

- All relevant new information is rapidly incorporated into the price
- Competition enables efficient allocation

How to improve judgments IV: Wisdom of the crowd x beliefs incentivization

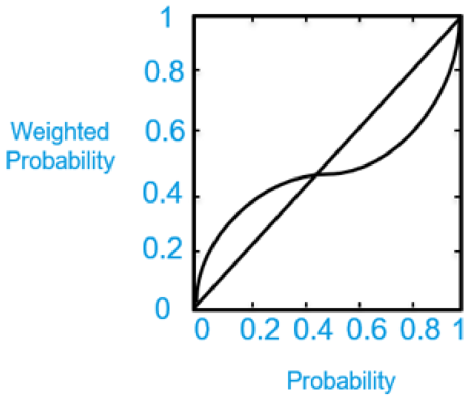
Prediction markets cleverly combine:

1. Wisdom of crowds

- Aggregated judgment from a large group is more accurate than that of a single expert
- Requires private, independent, and diverse predictions

2. Belief elicitation through monetary incentives (skin in the game) (Kant; Schotter et Trevino, 2014):

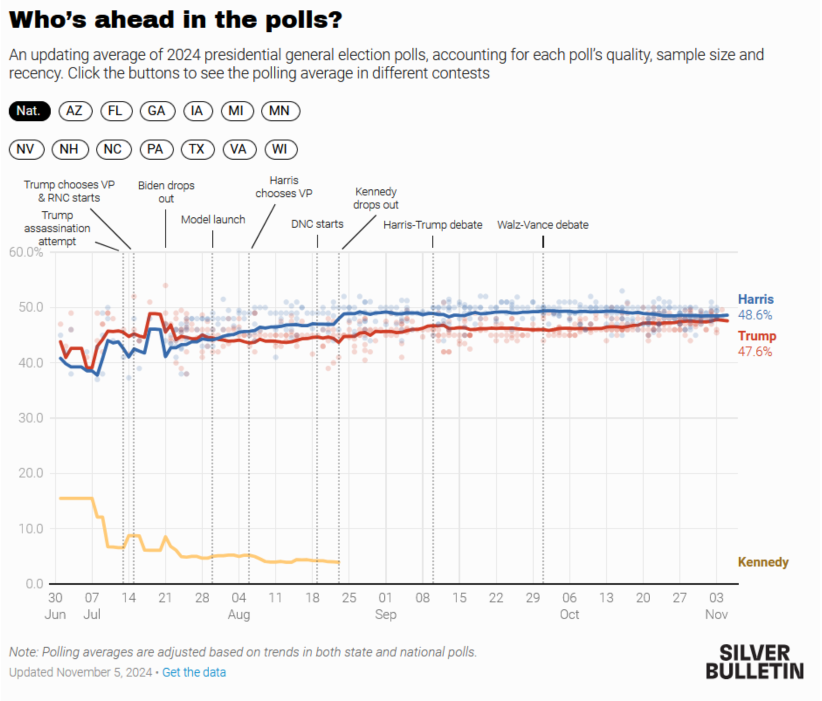
- Promotes accurate estimates and calibrated confidence

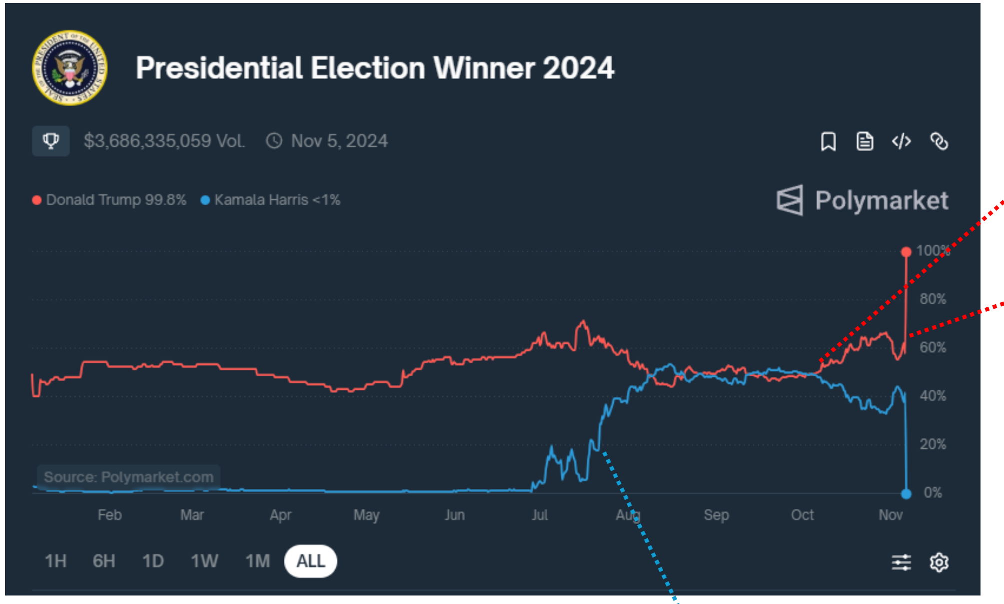


3. Efficient markets:

- All relevant new information is rapidly incorporated into the price
- Competition enables efficient allocation

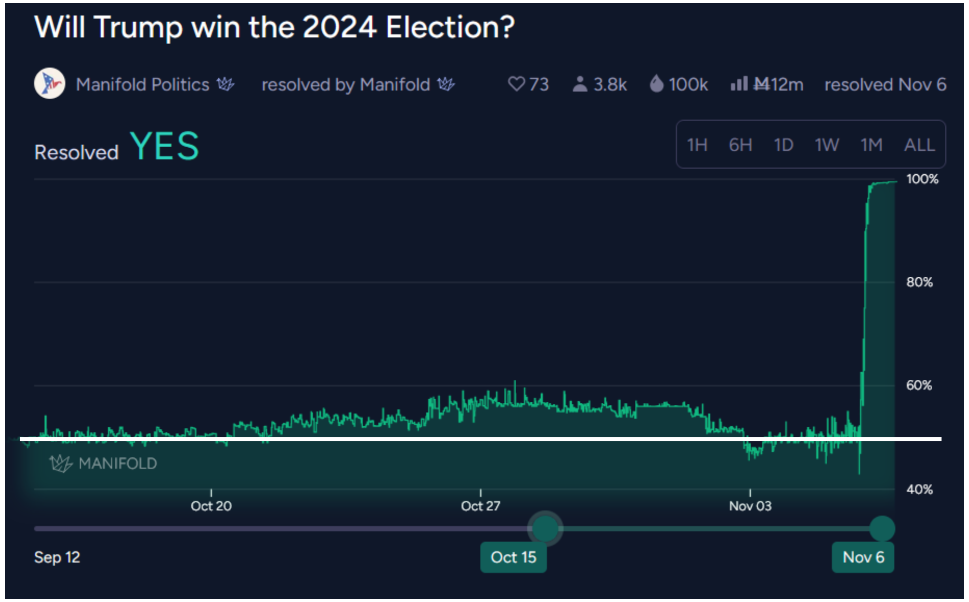
Crowds & incentivization: the case of prediction markets





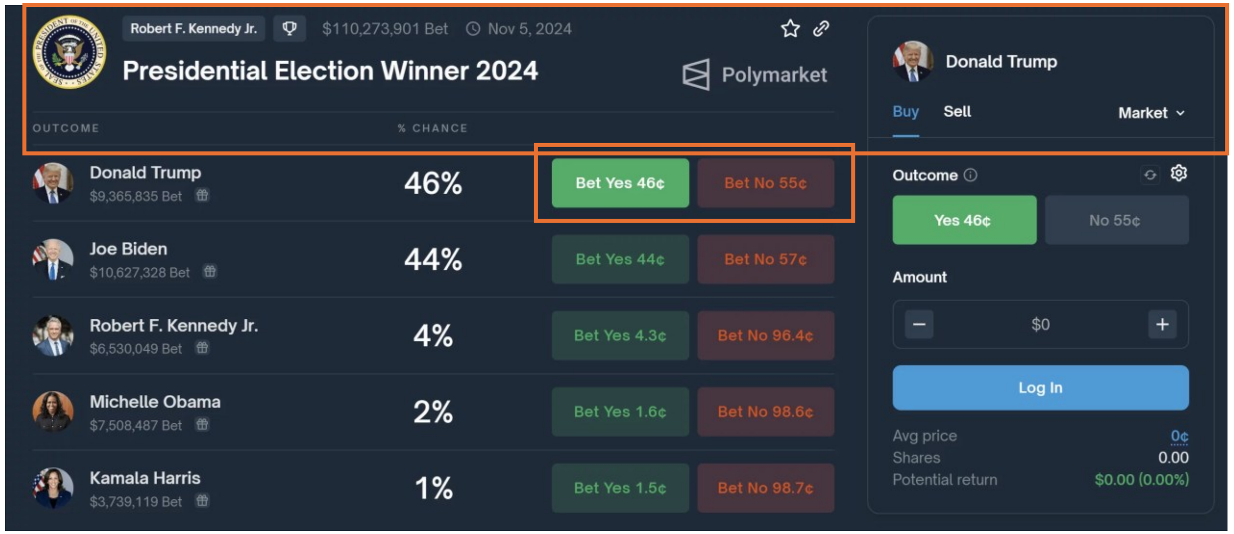
October 8

Novembrer 5
(day prior)



Kamala Harris announces her candidacy

Prediction markets: how they work



- Price of a share: probability as estimated by the market
- Favors relevant information
- Allows performance to be recorded and tracked
- Can be used to support decision-making (e.g., Google) or as a polling method

- Buying and selling shares representing the outcome of future events with predefined terms and conditions
- Two shares: YES and NO; each is priced between 0 and 1
- Match one buyer with one seller. Example: a buyer for YES at €0.57 and a buyer for NO at €0.43
- Resolution:
  - If the event occurs, each YES share is worth €1 and each NO share is worth €0
  - If the event does not occur, each NO share is worth €1 and each YES share is worth €0

Prediction markets: performances

4 Hours Before Markets Resolve Polymarket Accuracy %

94.3%
Accurate

@alexmccullough ··· 6d

12 Hours Before Markets Resolve Polymarket Accuracy %

90.5%
Accurate

@alexmccullough ··· 6d

1 Day Before Markets Resolve Polymarket Accuracy %

88.7%
Accurate

@alexmccullough ··· 6d

1 Week Before Markets Resolve Polymarket Accuracy %

89.4%
Accurate

@alexmccullough ··· 6d

1 Month Before Markets Resolve Polymarket Accuracy %

90.8%
Accurate

@alexmccullough ··· 6d

1 Hour After Start Time Polymarket Sporting Events Accuracy

80.7%
Accurate

@alex_m ··· 2d

Game Start Time Polymarket Sporting Events Accuracy

71.7%
Accurate

@alex_m ··· 2d

6 Hours Before Game Polymarket Sporting Events Accuracy

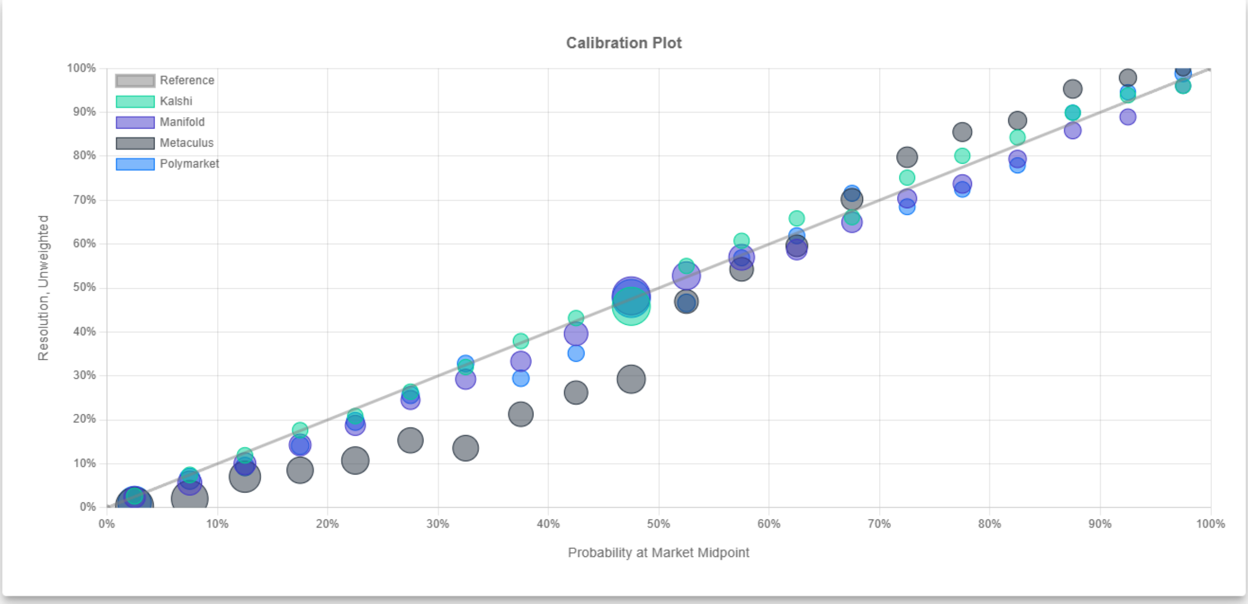
66.4%
Accurate

@alex_m ··· 2d

1 Day Before Game Polymarket Sporting Events Accuracy

66.1%
Accurate

@alex_m ··· 2d

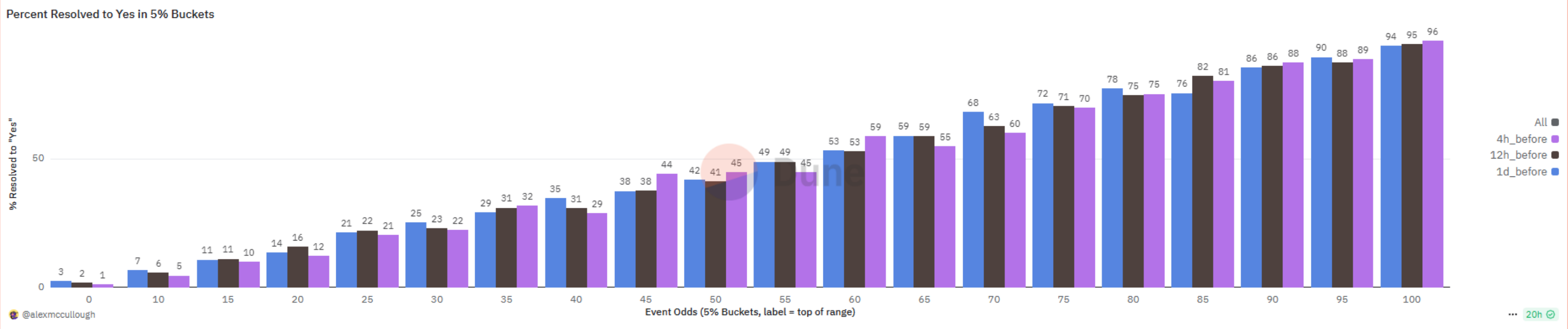


Kalshi
A US-regulated exchange with limited real-money contracts.

Manifold
A play-money platform where anyone can make any market.

Metaculus
A forecasting platform focused on calibration instead of bets.

Polymarket
A high-volume cryptocurrency exchange backed by USDC.



Prediction markets: methods to estimate plausibility

RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES |

Using prediction markets to estimate the reproducibility of scientific research

Anna Dreber, Thomas Pfeiffer, Johan Almenberg, +4, and Magnus Johannesson Authors Info & Affiliations

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved October 6, 2015 (received for review August 17, 2015)

November 9, 2015 | 112 (50) 15343-15347 | https://doi.org/10.1073/pnas.1516179112

VIEW RELATED CONTENT +

REPORT

Evaluating replicability of laboratory experiments in economics

COLIN F. CAMERER, ANNA DREBER, ESKIL FORSELL, TECK-HUA HO, JÜRGEN HUBER, MAGNUS JOHANNESSON, MICHAEL KIRCHLER, JOHAN ALMENBERG, ADAM ALTMEJD, [...], AND HANG WU +8 authors Authors Info & Affiliations

SCIENCE • 3 Mar 2016 • Vol 351, Issue 6280 • pp. 1433-1436 • DOI: 10.1126/science.aaf0918

JOURNAL ARTICLE | OPEN ACCESS | PEER REVIEWED

Predicting replication outcomes in the Many Labs 2 study

Eskil Forsell, Domenico Viganola, Thomas Pfeiffer, Johan Almenberg, Brad Wilson, Yiling Chen, Brian A. Nosek, Magnus Johannesson and Anna Dreber Show details for 9 authors

Journal of Economic Psychology, Vol.75(Part A SI), 102117
2019-12
DOI: https://doi.org/10.1016/j.joep.2018.10.009

Article | Published: 20 May 2020

Variability in the analysis of a single neuroimaging dataset by many teams

Rotem Botvinik-Nezer, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, ... Tom Schonberg + Show authors

Nature **582**, 84–88 (2020) | Cite this article

Letter | Published: 27 August 2018

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, Adam Altmejd, Nick Buttrick, Taizan Chan, Yiling Chen, Eskil Forsell, Anup Gampa, Emma Heikensten, Lily Hummer, Taisuke Imai, Siri Isaksson, Dylan Manfredi, Julia Rose, Eric-Jan Wagenmakers & Hang Wu

Nature Human Behaviour **2**, 637–644 (2018) | Cite this article

How to improve judgments V: Probabilistic reasoning and calibration

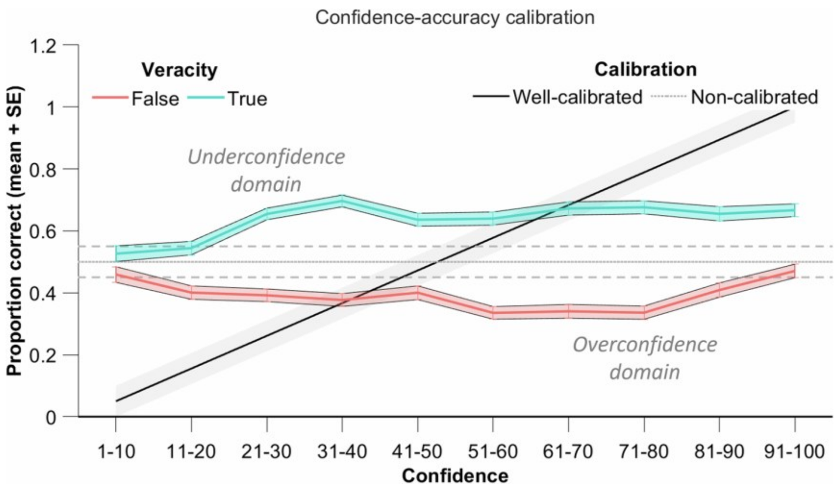
1. **Triage** (avoid wasting time on irrelevant problems)
2. **Break problems down**
3. **Balance inside and outside views** (identify comparison classes)
4. **Update beliefs** (Bayesian updating + calibrate confidence / base rates)
5. **Remain open to being wrong**
6. **Reduce uncertainty** (nuance matters; distinguish 60/40 from 55/45)
7. **Balance caution and decisiveness**
8. **Learn from failures and successes**
9. **Team management** (stepping back, precise questioning, constructive challenge)
10. **Balance opposing errors**

Philip Tetlock & Dan Gardner

Calibrating confidence

Calibration: ability to adjust one's confidence level according to uncertainty, quality of evidence, and stability of the context

- Confidence dissociated from actual knowledge:
  - In polarized domains (climate change, COVID-19)
    (Fischer & Fleming, 2024 ; Guigon, Villeval et Dreher, 2024)
  - In difficult tasks
    (Brewer & Wells, 2006 ; Moore & Healy, 2008)

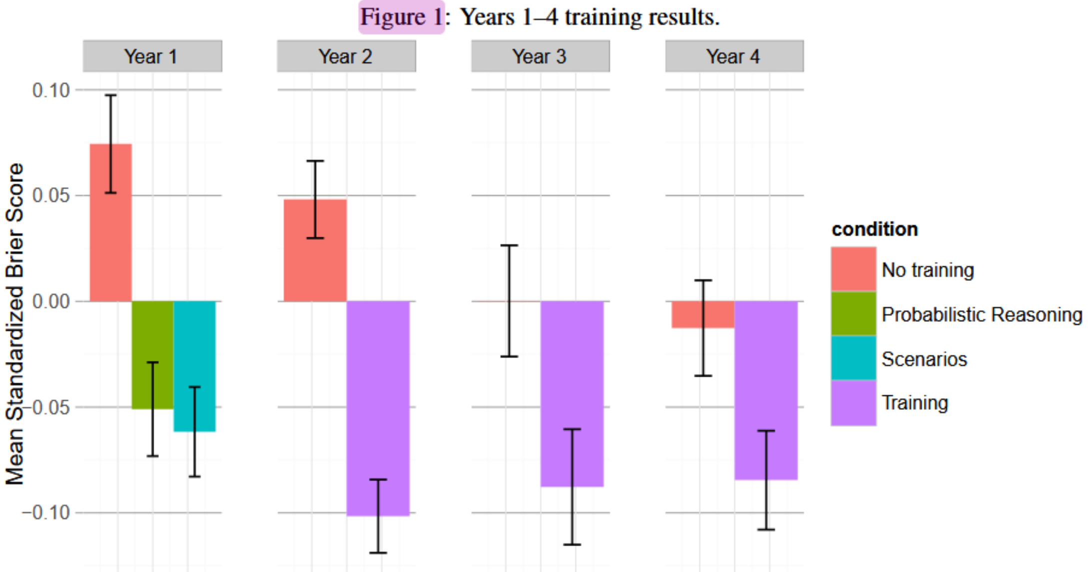


- Reducing overconfidence and improving judgment accuracy:
  - Individual feedback
  - Adaptive training
  - Digital tools for probabilistic estimation

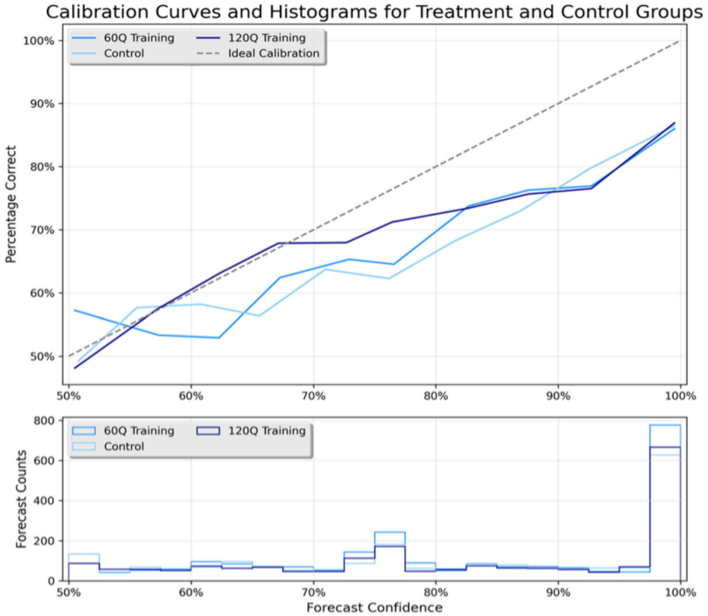
(Chang et al., 2016 ; Moore et al., 2017 ; Stone et al., 2023 ; Gruetzemacher et al., 2024 ; Motahhar et al., 2025)

- Metacognitive calibration promotes cognitive flexibility
- Metacognitive biases predict dogmatism and closed-mindedness
  (Rollwage et al., 2018 ; Fischer et al., 2019)

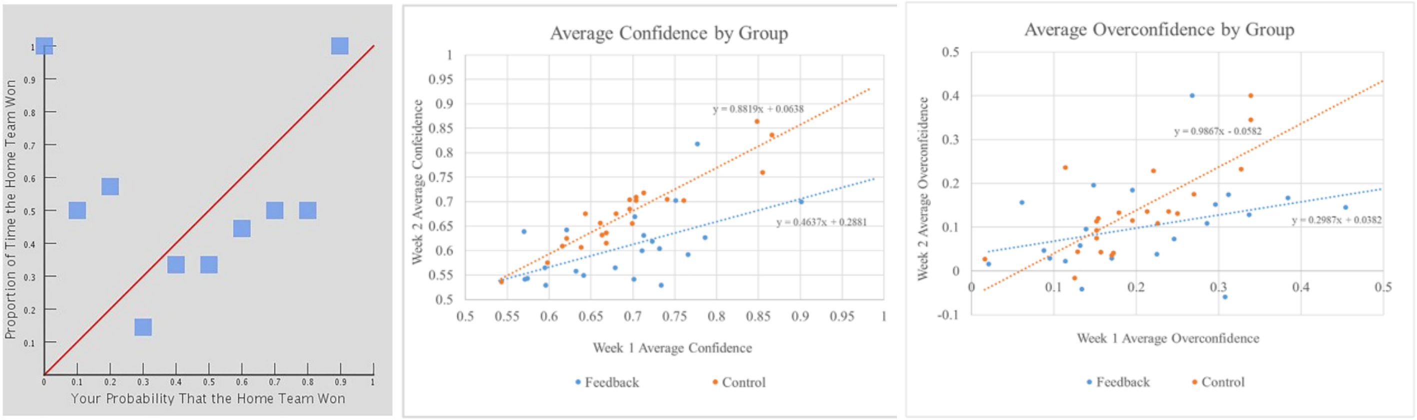
Training probabilistic reasoning and calibration



Chang et al, 2016 – Geopolitical forecasting
Recruitment: professionals, researchers, alumni associations, blogs, etc.
Training: reasoning principles and probabilistic reasoning



Gruetzemacher et al., 2024 – students majoring in business and sport culture: prediction of football game winners



Stone et al., 2023 – left: example of feedback; right: calibration curve
Recruitment: students interested in baseball

Probabilistic reasoning and calibration: the case of geopolitical forecasting

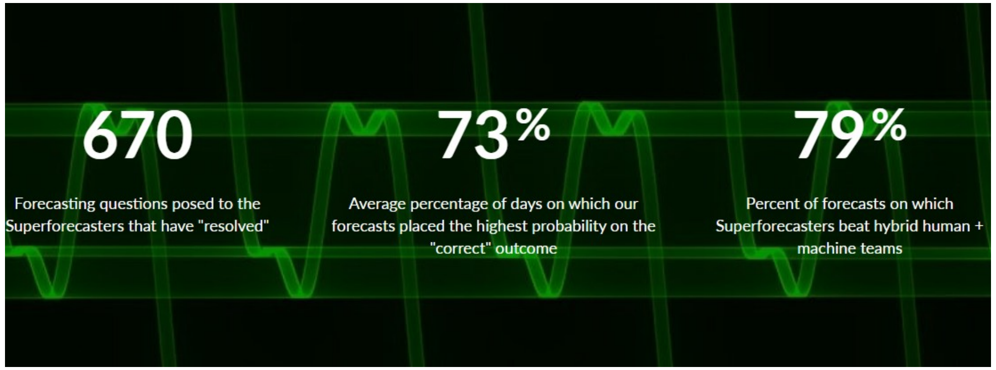
Announcement

Happy Friday, forecasters! We have nine new questions for your consideration:

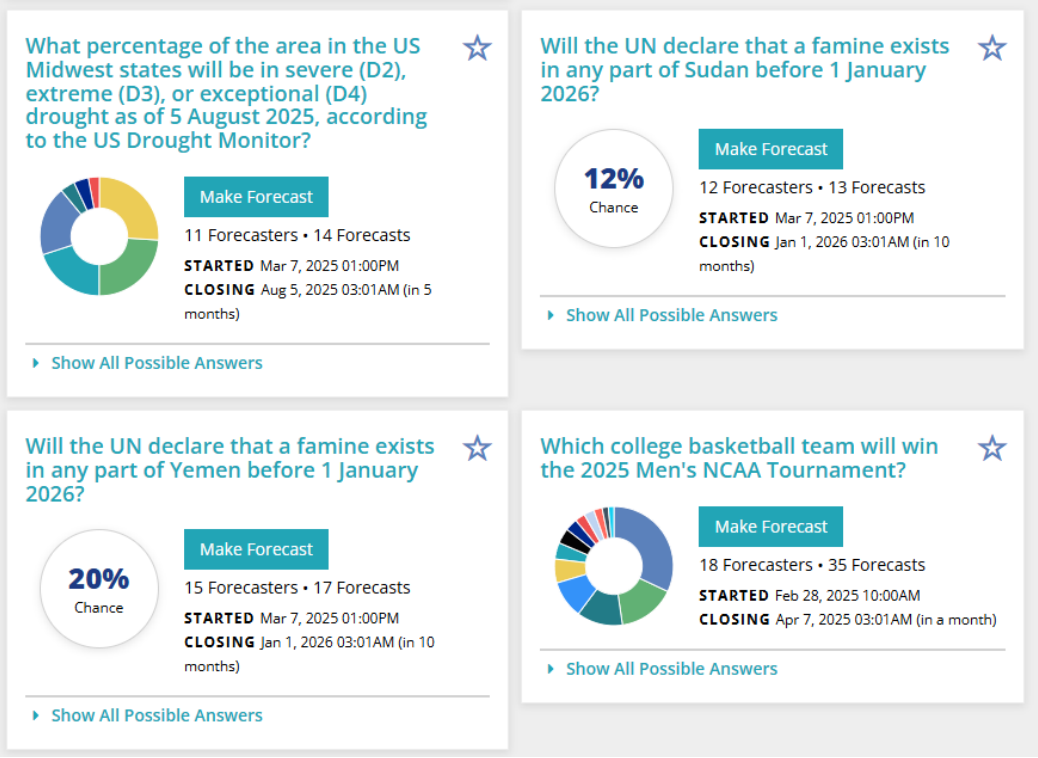
1. Will the UN declare that a famine exists in any part of Yemen before 1 January 2026?
2. Will the UN declare that a famine exists in any part of Sudan before 1 January 2026?
3. What percentage of the area in the US Midwest states will be in severe (D2), extreme (D3), or exceptional (D4) drought as of 5 August 2025, according to the US Drought Monitor?
4. How many total cases of dengue fever will the World Health Organization report in Brazil in the first half of 2025?
5. Before 1 January 2026, will Nancy Pelosi publicly announce that she will not run for reelection to the US House of Representatives in 2026?
6. Will the average daily crude oil production by Iran fall below 2,750 thousand barrels per day (tb/d) for any month in 2025?
7. What will be NVIDIA's total revenue in the first quarter of its fiscal year 2026 (approximately February through April 2025)?
8. Which football (soccer) club will win the 2024-25 Football Association Challenge Cup (FA Cup)?
9. Which football (soccer) club will win the 2024-25 Women's Football Association Challenge Cup (Women's FA Cup)?

Make your forecasts!

Mar 7, 2025 01:00PM

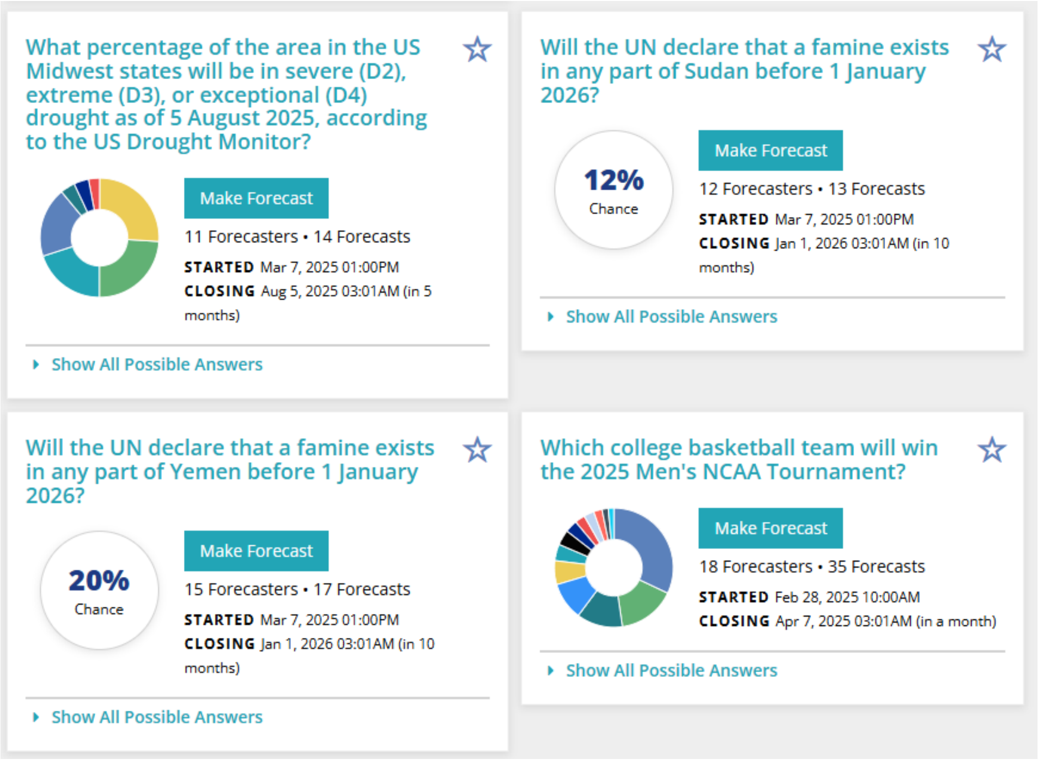


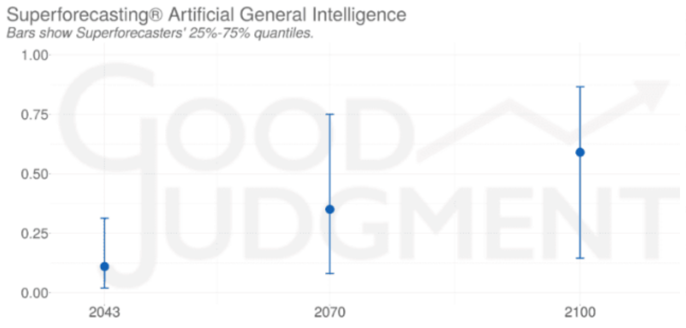
Good Judgment Open



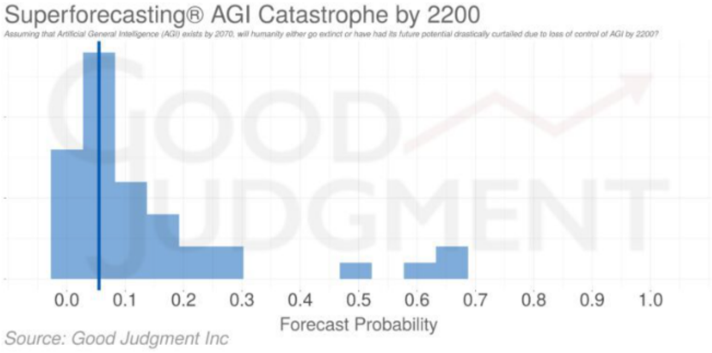
How to improve judgments VI:
Crowds x incentives x [proba. training + calibration]

Combine intellectual processes based on niche knowledge, probabilistic judgment, and performant crowds to estimate probabilities of **highly uncertain events**





Will AGI exist by 2043, 2070, or 2100? The median probabilities and 25%-75% quantiles as of 6 April 2023 sugge an increasing likelihood of AGI over the next 70 years with increasing variance/disagreement among Good Judgment's Superforecasters. (AGI, as defined in this project, could be said to exist if "for any human that can do a job, there is a computer program...that can do the same job for $25/hour or less." For a complete definition, pleas see the Supplementary Report.)



Assuming that AGI exists by 2070, will humanity either go extinct or have had its future potential drastically curtailed due to loss of control of AGI by 2200? Histogram of individual forecasts, with the dark blue line representing the median forecast.

Black swans (low probability & potentially massive impact)

- Limited historical data, high uncertainty, complex interdependencies
- Resistant to standard statistical approaches

(Atanasov et al., 2024; Karger et al., 2022)

Good Judgment .com & Good Judgment Open

How to improve judgments VII: Wisdom of human-silicon crowds

AI-Augmented Predictions: LLM Assistants Improve Human Forecasting Accuracy

PHILIPP SCHOENEGGER, LSE, London, United Kingdom of Great Britain and Northern Ireland
PETER S. PARK, MIT, Cambridge, MA, USA
EZRA KARGER, Federal Reserve Bank of Chicago, Chicago, IL, USA
SEAN TROTT, University of California San Diego, San Diego, CA, USA
PHILIP E. TETLOCK, University of Pennsylvania, Philadelphia, PA, USA

SCIENCE ADVANCES | RESEARCH ARTICLE

COMPUTER SCIENCE

Wisdom of the silicon crowd: LLM ensemble prediction capabilities rival human crowd accuracy

Philipp Schoenegger[1]*, Indre Tuminauskaite[2], Peter S. Park[3], Rafael Valdece Sousa Bastos[4], Philip E. Tetlock[5,6]