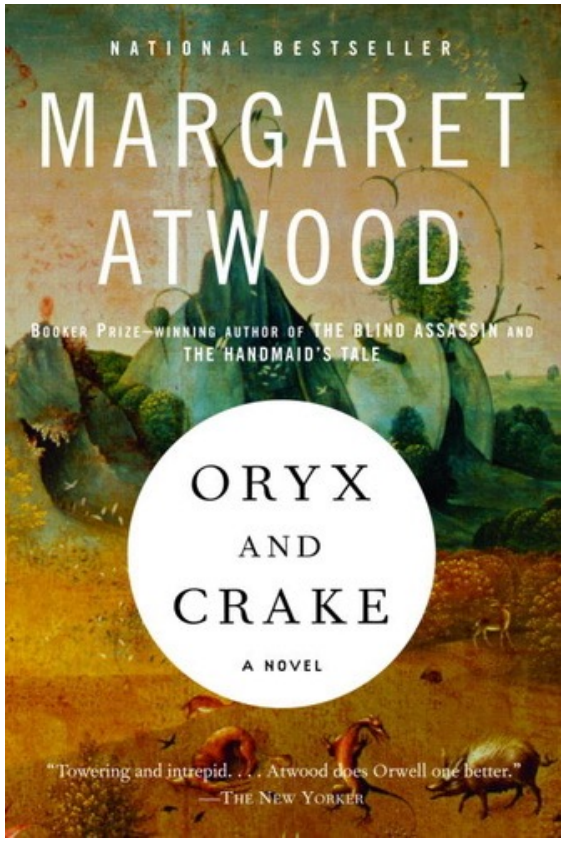
Kognitionspsychologie: Session 10

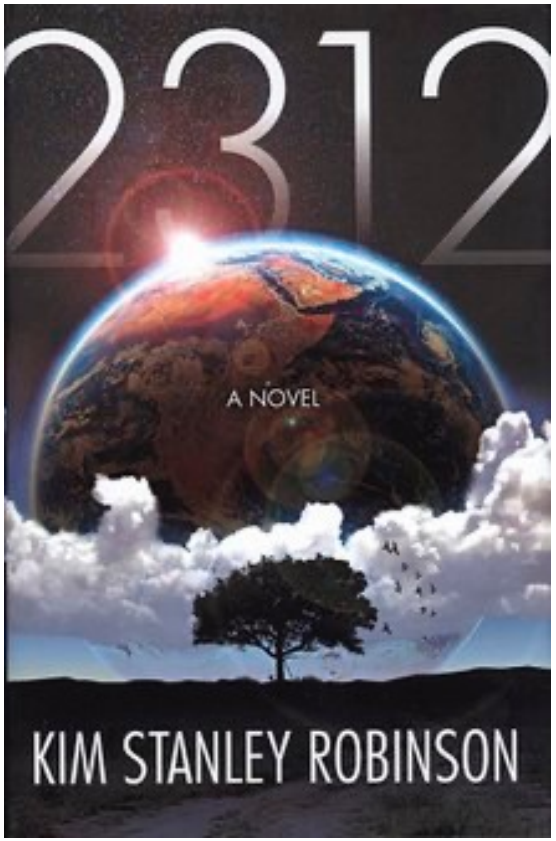
Wrap-up and Q&A

Rui Mata, HS 2024

Version: December 17, 2024

Science fiction as a tool…





"People sometimes think that science fiction is about predicting the future, but that isn't true. (…) science fiction is more of a modeling exercise, or a way of thinking."

Kim Stanley Robinson

Historical analysis as a tool…

YOU CAN'T PLAY 20 QUESTIONS WITH NATURE
AND WIN:
PROJECTIVE COMMENTS ON THE PAPERS OF THIS SYMPOSIUM

Allen Newell
May, 1973

I am a man who is half and half. Half of me is half distressed and half confused. Half of me is quite content and clear on where we are going.

My confused and distressed half has been roused by my assignment to comment on the papers of this symposium. It is curious that it should be so. We have just listened to a sample of the best work in current experimental psychology. For instance, the beautifully symmetric RT data of Cooper and Shepard (Chapter 3) make me positively envious. It is a pleasure to watch Dave Klahr (Chapter 1) clean up the subitizing data. The demonstrations of Bransford and Johnson (Chapter 8) produce a special sort of impact. And so it goes. Furthermore, independent of the particular papers presented here, the speakers constitute a large proportion of my all-time favorite experimenters--Chase, Clark, Posner, Shepard. Not only this, but almost all of the material shown here serves to further a view of man as a processor of information, agreeing with my current theoretical disposition. Half of me is ecstatic.

Still, I am distressed. I can illustrate it by the way I was going to start my comments, though I could not in fact bring myself to do so. I was going to draw a line on the blackboard and, picking one of the speakers of the day at random, note on the line the time at which he got his PhD and the current time (in mid-career). Then, taking his total production of papers like those in the present symposium, I was going to compute a rate of productivity of such excellent work.. Moving, finally, to the date of my chosen target's retirement, I was going to compute the total future additon of such papers to the (putative) end of this man's scientific career. Then I was going to pose, in my role as discussant, a question: Suppose you had all those additional papers, just like those of today (except being on new aspects of the problem), *where will psychology then be?* Will we have achieved a science of man adequate in power and commensurate with his complexity? And if so, how will this have happened via these papers that I have just granted you? Or will we be asking for yet another quota of papers in the next dollop of time?

Some of Newell's solutions:

- *Analyze complex tasks*
- *Create complete processing models (e.g., from perception to action)*
- *Address multiple levels of analysis and adopt interdisciplinary perspective*

A few questions about KOGPSY in 2050…

1. Will we still need pluralistic explanations?
2. Will our models of intelligence involve *g*? And, if so, how?
3. Will we (still) think of the mind as a collection of modules?
4. Will machines have consciousness?
5. Will behavioral research still be needed?

Will we still need pluralistic explanations?

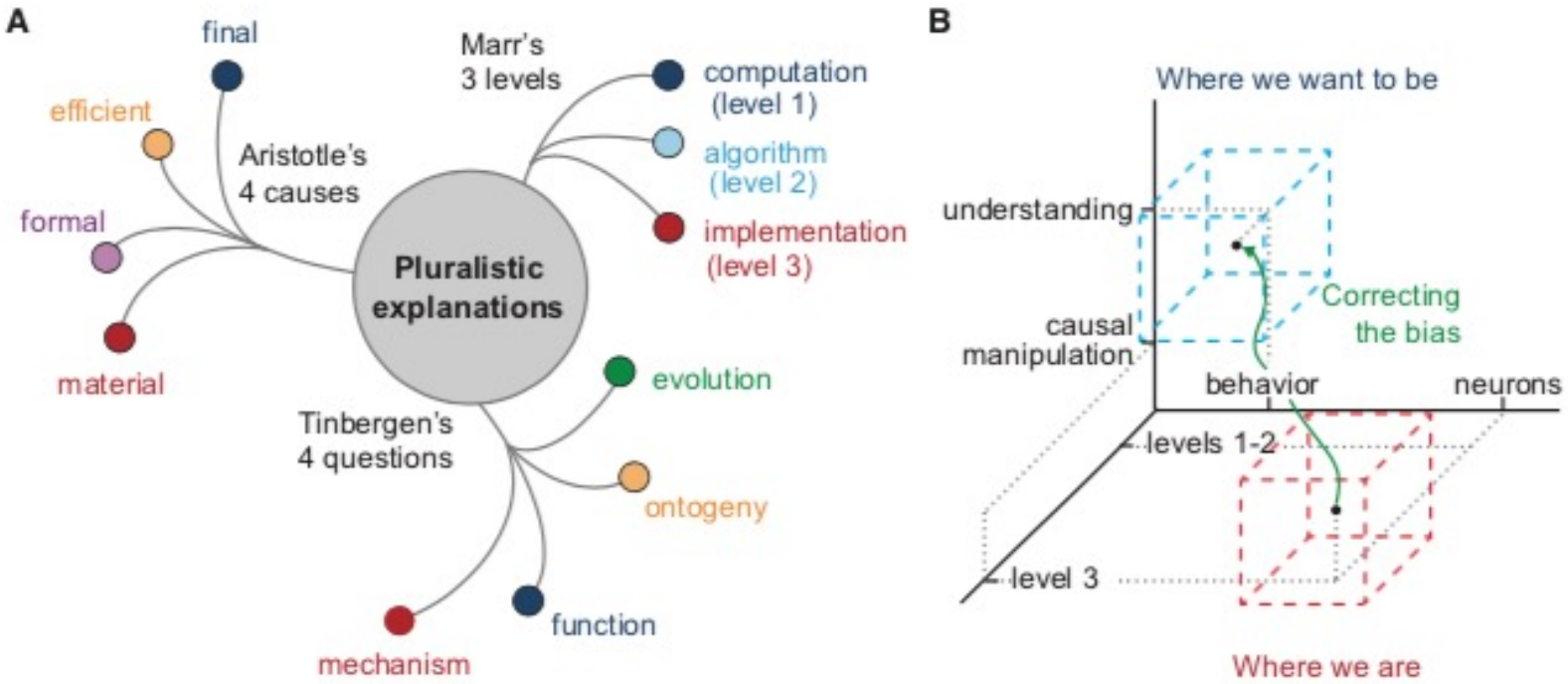


Figure 4. The Future History of Pluralistic Explanation

(A) That understanding of a phenomenon is multidimensional has long been appreciated. Aristotle posited four kinds of explanation: to explain "why" something changes, a polyhedric notion of causality is necessary; one that includes not only the material cause (what it is made out of), but also the other three "whys": formal (what it is to be), efficient (what produces it), and final (what it is for). Tinbergen also devised four questions about behavior: to go beyond its proximate causation (mechanism) to also considering its evolution, development, and real-world function. Marr's three levels are also shown.

(B) Three-dimensional space with axes of understanding-manipulation, behavior-neurons, and Marr's levels. The red box is where we are and the blue is where we should be.

Will our models of intelligence involve *g*?

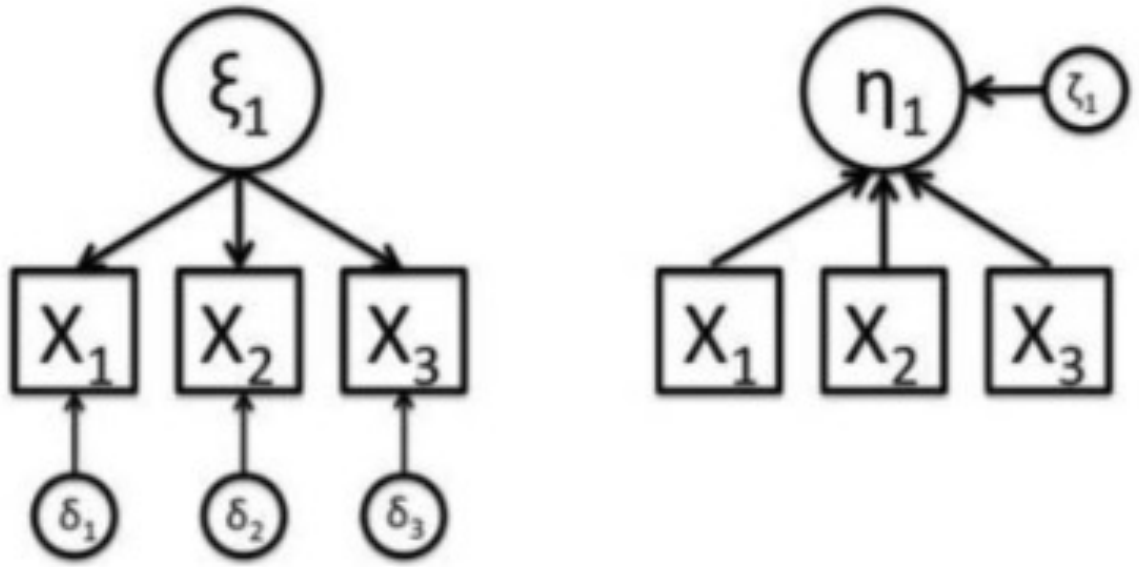
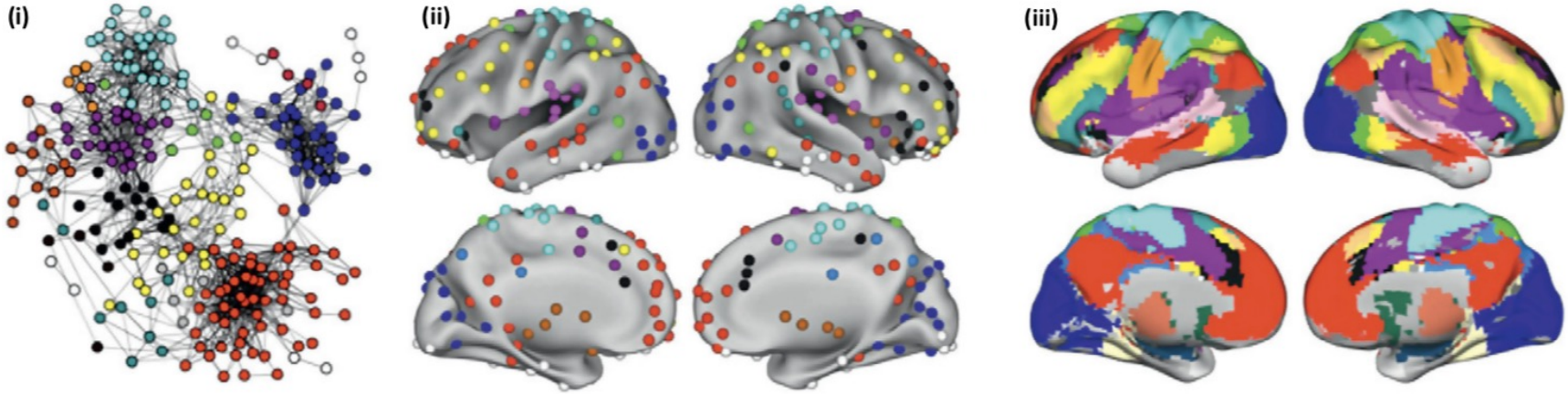


Figure 7. A reflective (left) and a formative (right) model.

Will we think of the mind as a collection of modules?



Will machines have consciousness?

Can machines have consciousness?

Type	Description	Example(s)	
C0: Unconscious processing	Information processing can be realized by (mindless) automatons	face or speech recognition, priming, debating(!)	Yes
C1: Global availability	Selection of information for global broadcasting, making it robust, and available for computation and report	reportable aspects of sensory experience	Not yet
C2: Self-monitoring	Self-monitoring of computations, leading to a subjective sense of certainty or error.	confidence, error-monitoring, knowledge of strategy efficacy	

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? ***Science***, 358(6362), 486–492. http://doi.org/10.1126/science.aan8871

7

POL CARTE/SIPA/ALAMY

Conscious AI systems could suffer if people neglect them or treat them poorly.

WHAT HAPPENS IF AI BECOMES CONSCIOUS? IT'S TIME TO PLAN

Tech companies urged to test systems for capacity for subjective experience, and make policies to avoid harm.

By Mariana Lenharo

The rapid evolution of artificial intelligence (AI) is bringing up ethical questions that were once confined to science fiction: if AI systems could one day 'think' like humans, for example, would they also be able to have subjective experiences like humans? Would they experience suffering, and, if so, would humanity be equipped to care for them properly?

A group of philosophers and computer scientists is arguing that AI welfare should be taken seriously. In a report posted last month on the preprint server arXiv, ahead of peer review, the group calls for AI companies not only to assess their systems for evidence of consciousness and the capacity to make autonomous decisions, but also to put in place policies for how to treat the systems if

The stakes are getting higher as we become increasingly dependent on these technologies, says Jonathan Mason, a mathematician based in Oxford, UK. Mason argues that developing methods to assess AI systems for consciousness should be a priority. "It wouldn't be sensible to get society to invest so much in something and become so reliant on something that we knew so little about – that we didn't even realize that it had perception," he says.

People might also be harmed if AI systems aren't tested properly for consciousness, says Jeff Sebo, a philosopher at New York University in New York City and a co-author of the report. If we wrongly assume a system is conscious, he says, welfare funding might be funnelled towards its care, and therefore taken away from people or animals that need it. Furthermore, "it could lead you to constrain efforts to make AI safe or beneficial for humans".

A turning point?

The report contends that AI welfare is at a "transitional moment". One of its authors, Kyle Fish, was recently hired as an AI-welfare researcher by the AI firm Anthropic, based in San Francisco, California. This is the first such position of its kind designated at a top AI firm, according to authors of the report. Anthropic also helped to fund initial research that led to the report. "There is a shift happening because there are now people at leading AI companies who take AI consciousness and agency and moral significance seriously," Sebo says.

Nature contacted four leading AI firms to ask about their plans for AI welfare. Three – Anthropic, Google and Microsoft – declined to comment, and OpenAI, based in San Francisco, did not respond.

Some are yet to be convinced that AI consciousness should be a priority. In September, the United Nations High-level Advisory Body

Will behavioral research still be needed?

The Primacy of Behavioral Research for Understanding the Brain

Yael Niv
Department of Psychology and Neuroscience Institute, Princeton University

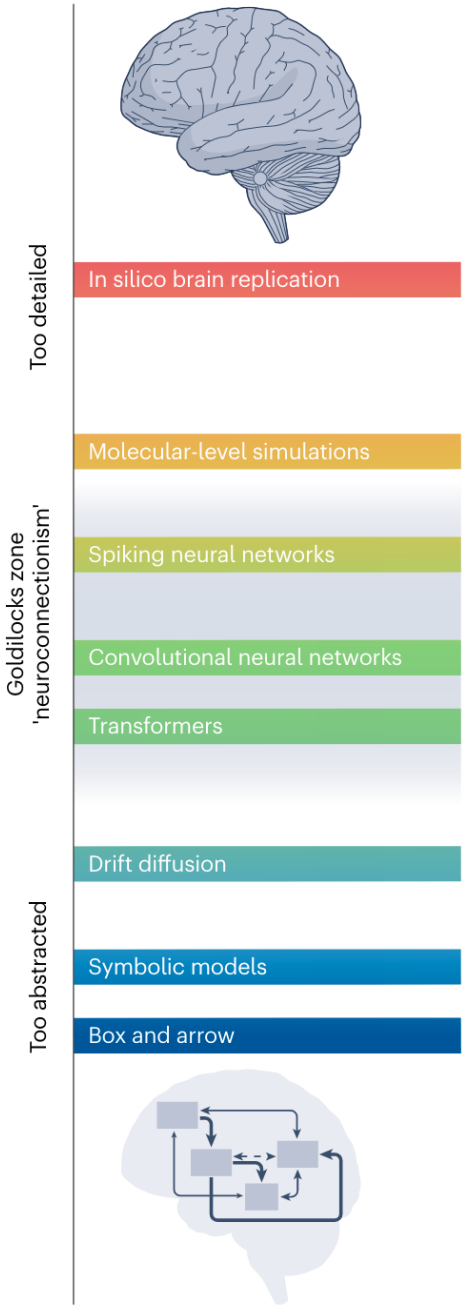
Understanding the brain requires us to answer both *what* the brain does, and *how* it does it. Using a series of examples, I make the case that behavior is often more useful than neuroscientific measurements for answering the first question.[1] Moreover, I show that even for "how" questions that pertain to neural mechanism, a well-crafted behavioral paradigm can offer deeper insight and stronger constraints on computational and mechanistic models than do many highly challenging (and very expensive) neural studies. I conclude that purely behavioral research is essential for understanding the brain—especially its cognitive functions—contrary to the opinion of prominent funding bodies and some scientific journals, who erroneously place neural data on a pedestal and consider behavior to be subsidiary.

Keywords: behavioral experiments, cognition, neuroscience, priorities

Niv, Y. (2021). The primacy of behavioral research for understanding the brain. *Behavioral Neuroscience, 135*(5), 601–609. https://doi.org/10.1037/bne0000471

What did we NOT cover?

In our course, we often discussed models that may be considered “too abstracted” (i.e., symbolic, box and arrow models). However, more detailed theories and models already exist and future theorizing is likely to become even more detailed/concrete and based on neurocomputational principles (cf. Doerig et al., 2023).



Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., Van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7), 431–450. https://doi.org/10.1038/s41583-023-00705-w

What did we NOT cover?

In our course, we covered a small set of empirical findings often relying on "traditional" methods (e.g., behavioral experiments, lesions) but ignored many others or covered them only briefly (e.g., fMRI, single-unit recording). A complete understanding of cognition will likely require many different (and ideally) converging methods with different strengths and weaknesses.

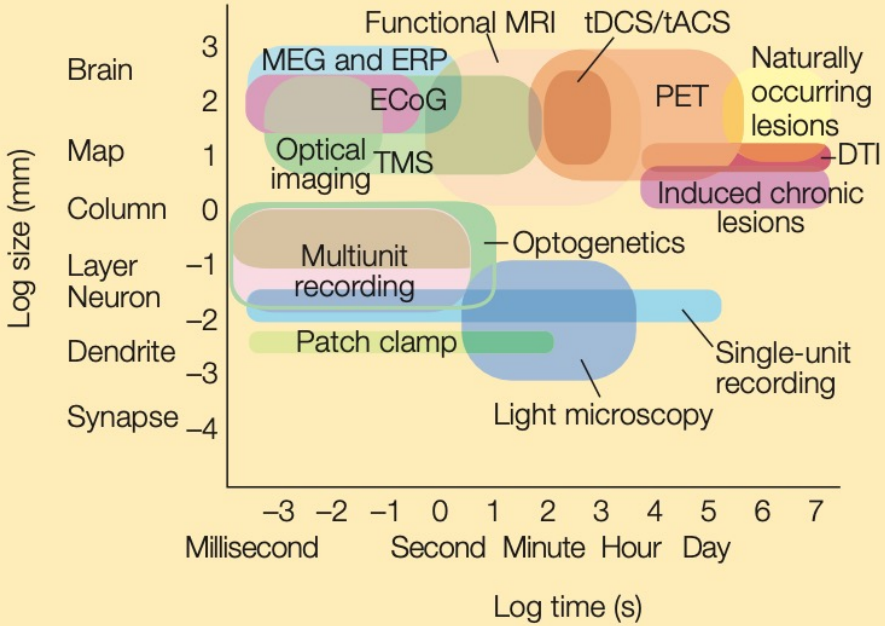
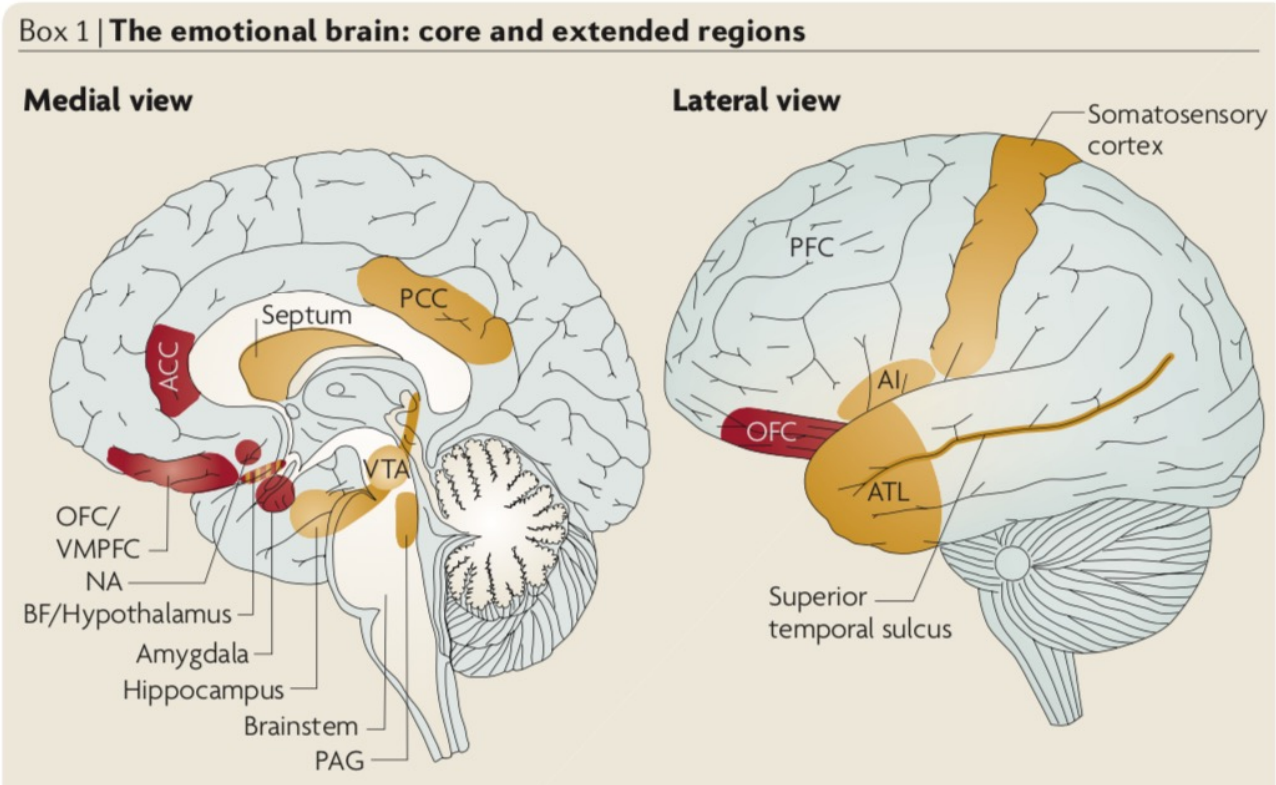


FIGURE 3.45 Spatial and temporal resolution of the prominent methods used in cognitive neuroscience.
Temporal sensitivity, plotted on the *x*-axis, refers to the timescale over which a particular measurement is obtained. It can range from the millisecond activity of single cells to the behavioral changes observed over years in patients who have had strokes. Spatial sensitivity, plotted on the *y*-axis, refers to the localization capability of the methods. For example, real-time changes in the membrane potential of isolated dendritic regions can be detected with patch clamps, providing excellent temporal and spatial resolution. In contrast, naturally occurring lesions damage large regions of the cortex and are detectable with MRI.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2018). Cognitive neuroscience: The biology of the mind (5th ed.). W.W. Norton & Company.

What did we NOT cover?

In our course, we largely ignored motivational and emotional aspects. However, these dimensions are crucial to modern theories of cognition and decision-making. We will focus on these aspects in KOGPSY II…



Next weeks

- You can submit final questions through ADAM until **January 15th**