Debate #3
Does free will exist?
What is the relationship between determinism and causality?

Debate panel: *Shinsuke Shimojo / Hakwan Lau / Hayato Saigo*

Will neural determinism exorcise “free will”?

- A new approach from psychophysical compatibilism -

Shinsuke Shimojo

Division of Biology & Biological Engineering / Computation and Neural Systems,
California Institute of Technology (Pasadena, CA).

NIH, NSF, HFSP.
JST.ERATO Shimojo "Implicit Brain Function" Project
MIXT Tamagawa-Caltech GCOE.
JST.CREST "Implicit InterPersonal Information" Project
JST.CREST “Implicit Ambient Surface Information” Project

Central Dogma in Neuroscience

All mental experiences/phenomena have ***neural correlates*** (***neural causes***).

* Even "voluntary will" is ***not*** an exception.
* Neural correlates not only precede mental experiences, but ***actually cause*** them.

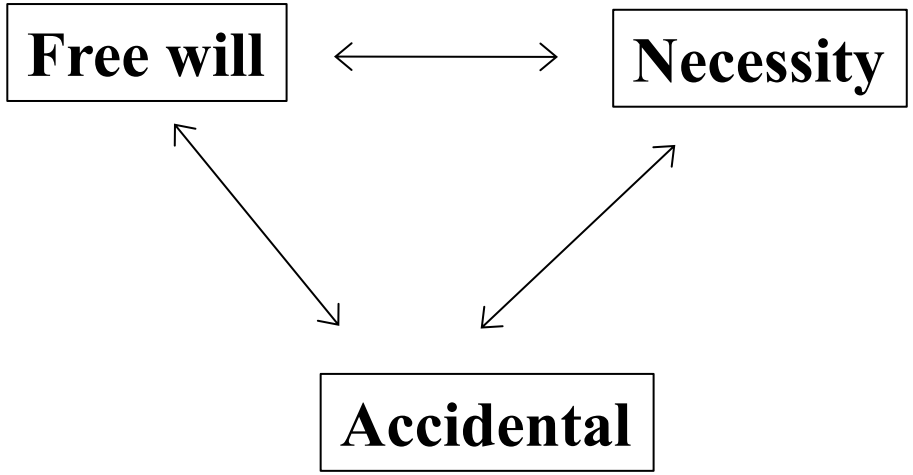
⇩

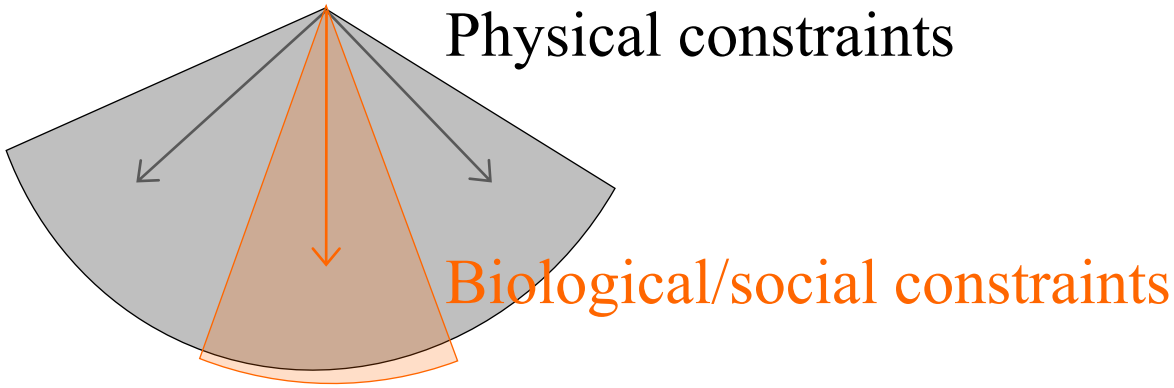
If causally determined by preceding neural processes, then no longer "free"?

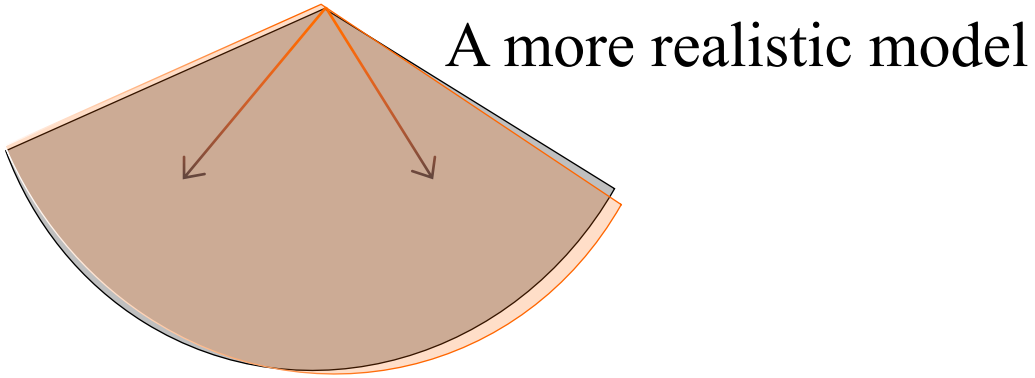
→ ***Neuroscientific determinism/reductionism***.

→ An ethical crisis of the social system, related to guilt, punishment & responsibility?

Can the neurosci. dogma & our sense of ***independency & freedom*** be compatible?







Spontaneous/Voluntary = Purposive

Spontaneous/Voluntary

Biological/social preceded?

* Various philosophical attempts to save free will from physical/neuroal determinism (compatibilism; dualism) → not very successful.

* The modern concept of "freedom" has only a short history (18th-19th century～).
It was ***retroactively constructed*** from the legal necessity of "accountability(responsibility)" (Kozakai, 2008; Shimojo, 2015).

← Necessity to draw a fine line between accountable vs. not-accountable cases

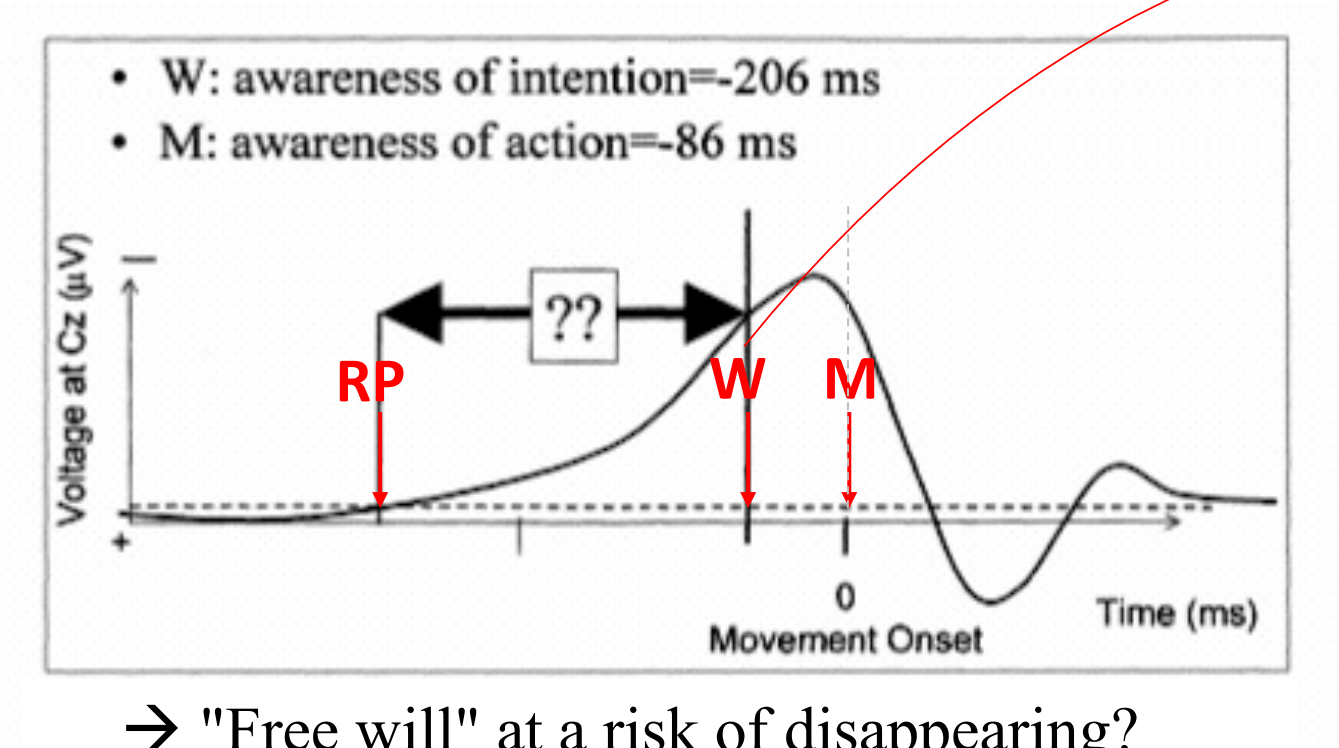
cf. non compos mentis (心神喪失)、feeble-mindedness (心神耗弱).

The free will concept has been developed just as ***a predominant fiction*** in criminal legal theories.

Neural activity precedes, and causes “free action”.

RP (Readiness Potential) 運動準備電位

：A slow upraising of electric activity in the motor cortex. Preceding to intended action ≥1sec. (Kornhuber & Deecke, 1965)



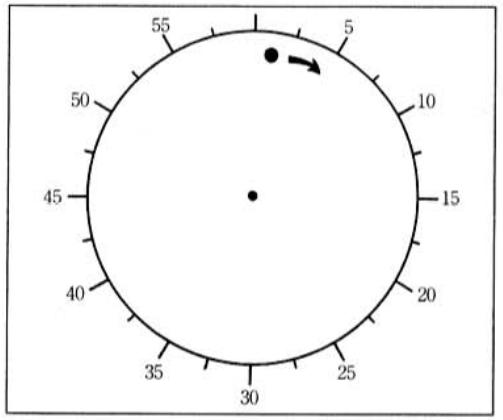


図 4.1　精神事象のタイミングを測る「時計」

（1 cycle／2.56 s）

(B. Libet)

→ "Free will" at a risk of disappearing?

Neural → Intention(will) → Action

Results of Libet et al. (1983)

RPI: Readiness potentials in trials where the action was planned within a crude range of timing. Peaked in the range from -800 to -1000ms（towards the onset of muscular activity）.

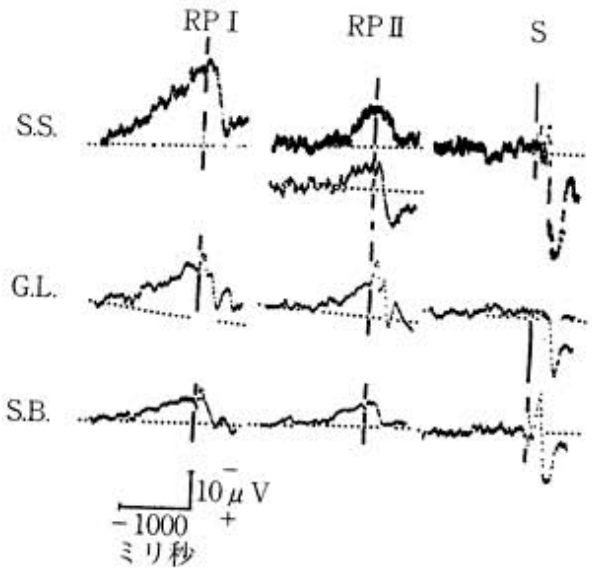


図 4.2　自発的に起動する自発的な行為に先行する準備電位（RP）

RPII: Rediness potentials in trials where the action was NOT planned at a particular timing. Peaked around -550ms.

(Libet 2004)

-550ms (RPII) –(- 150ms) (subjective onset of will) = **-400ms**.

Neural activity already starts, preceding 400 ms before the conscious awareness of will !

Fast vs. Slow W judgment: in relation to LRP

(An extension of B. Libet's)

	Early W trials	Late W trials
Mean W	–530	–179
LRP onset	–906	–713

Table 1. Mean LRP onset times and mean W (awareness of intention) judgements for trials showing early and late W. Data from Haggard and Eimer (1999).

LRPs, rather than RPs, are correlated better with W judgment in timing. The LRP precedes "W" by 380-530 ms.

Causal relationship ; Neural activity → Emergence of the "will"(?).

Even the most free & spontaneous decision is preceded by neural correlating activity.

It seemingly argues for ***Neural Determinism***, thus seemingly depriving freedom from the human. Yet, the human has a freedom to suppress its urge to act, and indeed has sufficient time (100s of ms) to do so (B.Libet. ***Mind Time***).

← Libet tried hard to rescue free will from his own neuroscientific findings, only to miserably fail.

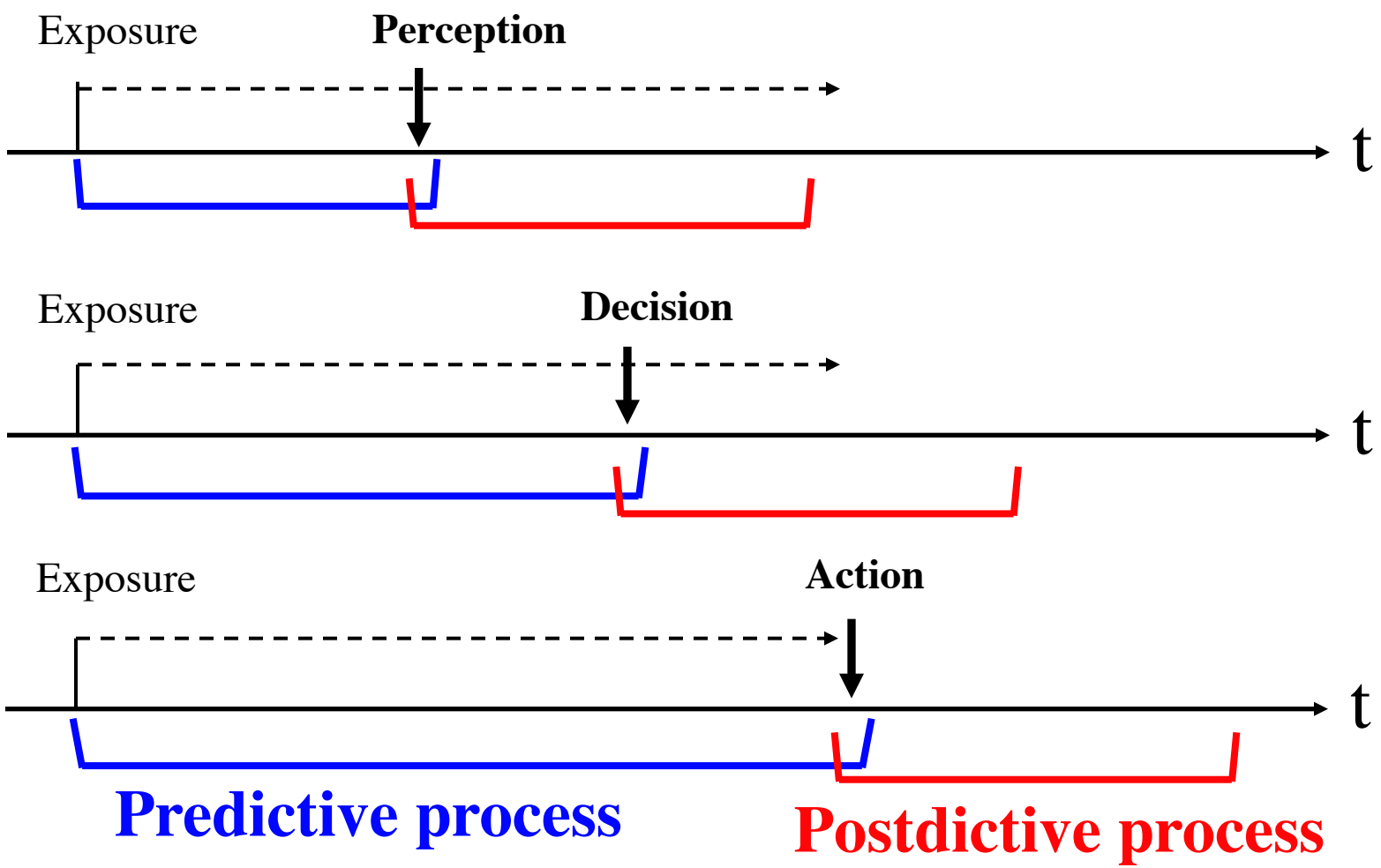
Three reasons why neural reductionism won't eliminate the sense of agency

(Shimojo fpsyg'14)

1. The feeling of free choice may live in the ***postdictive*** process, not in the predictive process.

cf. Cognitive dissonance (Festinger, '57), Causal attribution (Heider, '58), Choice justification (Staw, '76) - ***all postdictive!*** ***後付け再構成***

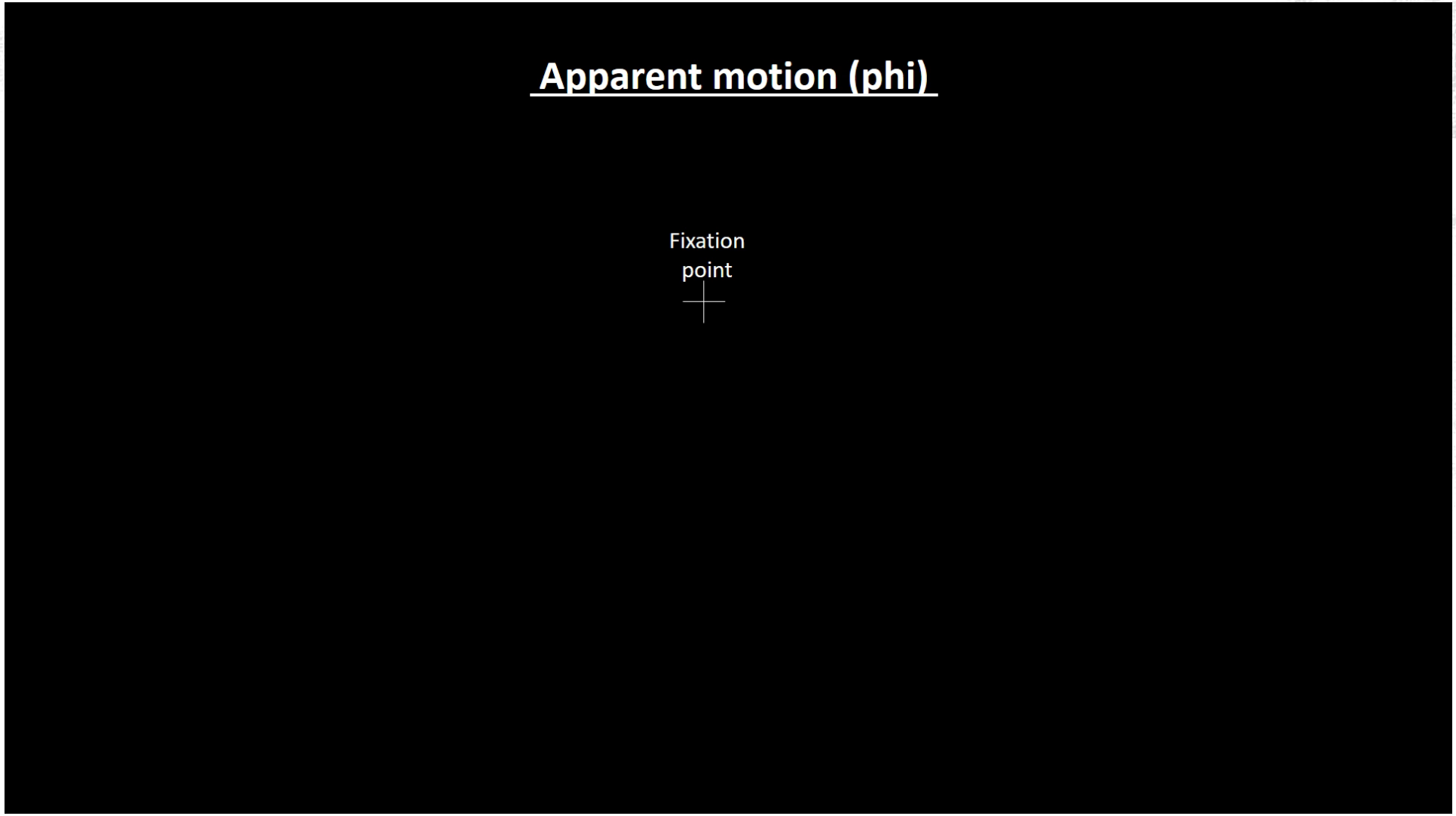
Prediction and Postdiction



To fully understand a mental experience of a moment, somatic and neural processing *before* ***& after*** may need to be considered (Shimojo, *SfN '16* plenary lecture) .

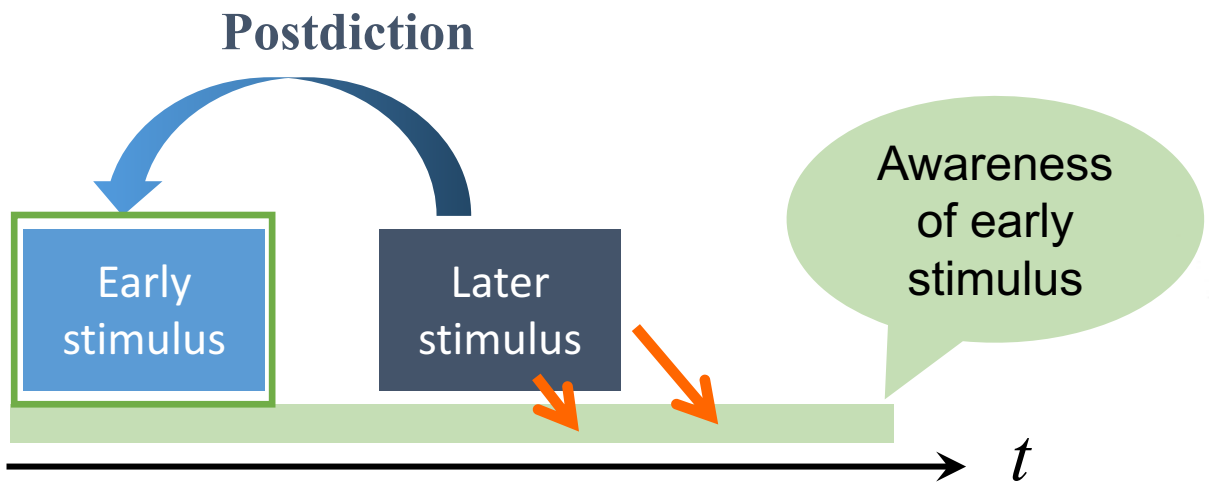
Apparent Motion

A classical example of perceptual postdiction (The “Phi Phenomenon”, M. Wertheimer, 1912)



“Postdictive” Phenomena

Perceptual (mental) phenomena, in which a stimulus presented later (in physical time) causally affects the perception of targets presented earlier (will extend later).



Initially proposed to account for the flash-lag effects, but actually has much broader applications / implications.

(Eagleman and Sejnowski, ‘00; Shimojo, '14; Yamada, *et al.*, '15)

Perceptual Postdiction

- Apparent motion
- Backward masking
- Flash lag (Nijihawan et al.)
- Double flash, AV rabbit (Shimojo et al.)

Cognitive Postdiction

- Constant updating of STM
- Causal (mis)attribution
- Memory distortion (cf. witness)
- Hindsight (Wu, et al.)

Choice Blindness 選択盲

Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task

Petter Johansson,[1*] Lars Hall,[1*†] Sverker Sikström,[1]
Andreas Olsson[2]

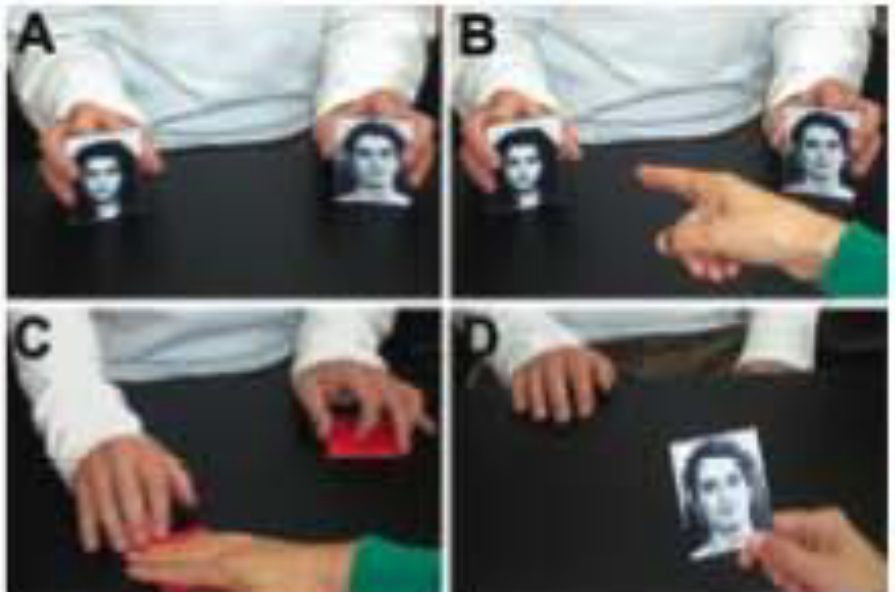
A fundamental assumption of theories of decision-making is that we detect mismatches between intention and outcome, adjust our behavior in the face of error, and adapt to changing circumstances. Is this always the case? We investigated the relation between intention, choice, and introspection. Participants made choices between presented face pairs on the basis of attractiveness, while we covertly manipulated the relationship between choice and outcome that they experienced. Participants failed to notice conspicuous mismatches between their intended choice and the outcome they were presented with, while nevertheless offering introspectively derived reasons for why they chose the way they did. We call this effect choice blindness.

(Science, 2005)

Choice Blindness

When presented the un-chosen face as just chosen, participants often did not notice & fluently describe the reason why they chose (the not-chosen).

Or, sometimes describe the real reason which is inconsistent (w. the presented face).



Choice Blindness : ***Postdictive*** *& explicit* reasoning (*cf.* Nisbett & Wilson, Psychol. Rev., '77).

Consciousness = a postdictive(, and upon-necessity) reconstruction.

eg. causal (mis)attribution, choice, justification, etc.

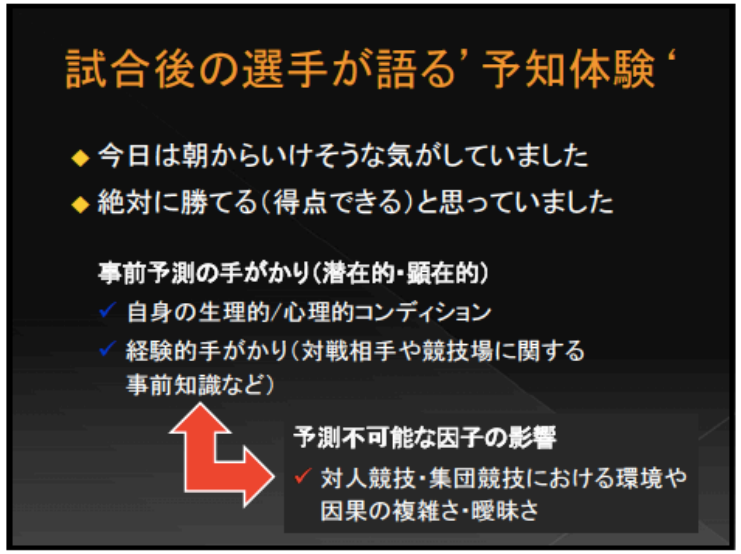
Athletes' "Sixth Sense"?

"I knew I could make a big splash in the game today."

← Is this real intuitive prediction, or rather a postdictive construct?

It turned out to be mostly postdictive, *i.e.* athletes modify their memory of prediction postdictively upon the outcome.

(kadota, Okumura & Shimojo, 2009)



Backward referral in reflexive response (B. Libet)

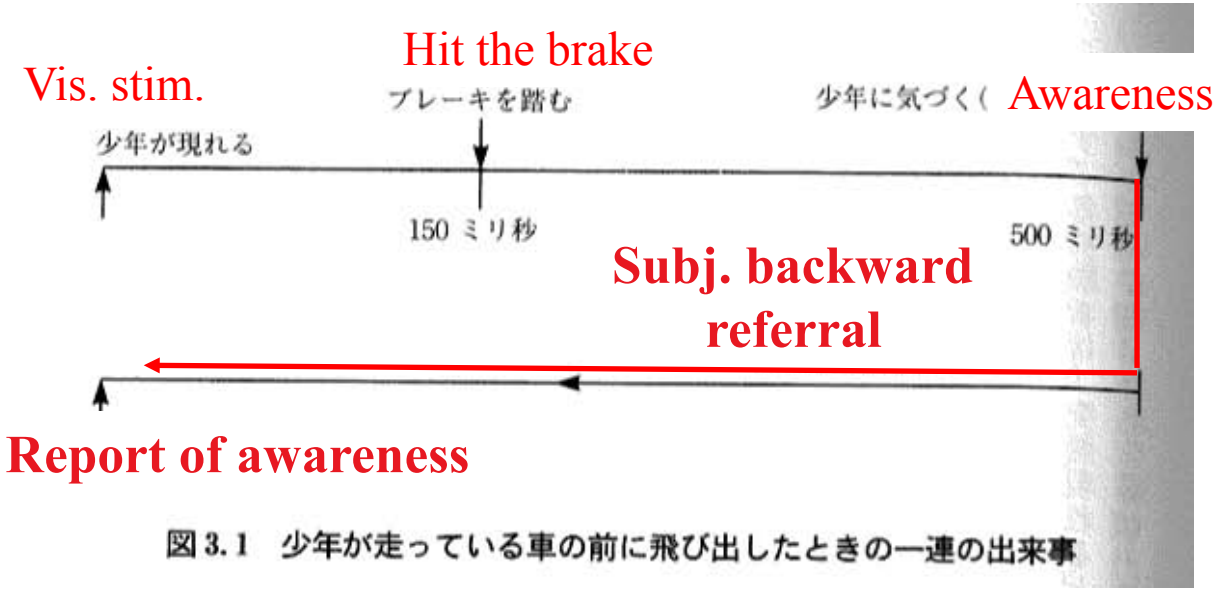


図 3.1　少年が走っている車の前に飛び出したときの一連の出来事

Reflexive response can be executed in 150ms.
Awareness (of the visual stimulus) requires additional 300-400ms,
but ***not*** perceived as ”delayed” from the response.
Rather, ***referred backwards*** to the moment before the response.

- Perception as revised history（Kinsbourne & Dennet 1992）
- A human = a quick “zombie” + a slow analyzer/speculator.

Mis-perceiving a TMS-triggered finger movement as a volitional action, indicating attribution to "volition." 意図の帰属

TMS applied to:	Hand selected Left	Right
Right M1	1029	771
Left M1	649	1155

# of trials (Ammon & Gandcvia, 1990)

(a) TMS biases the "volitional" selection only when RT<200ms (Brasil-Nato, *et al*., 1992).

(b) Subjects were unaware of the TMS effect.

<---> Readiness potential (Libet, et al., 1983) ?
Coupling between "consciousness" and "free will"

cf. Maeda, Wu & Shimojo (*unp.*): Role of the somatosensory cortex in perception of volitional movements.

Implicit motor command → Motor cortex activity → Muscle movement → Perception of intention (***postdictive attribution*** to "intention")

Three reasons why neural reductionism won't eliminate the sense of agency

(Shimojo fpsyg'14)

1. The feeling of free choice may live in the ***postdictive*** process, not in the predictive process.

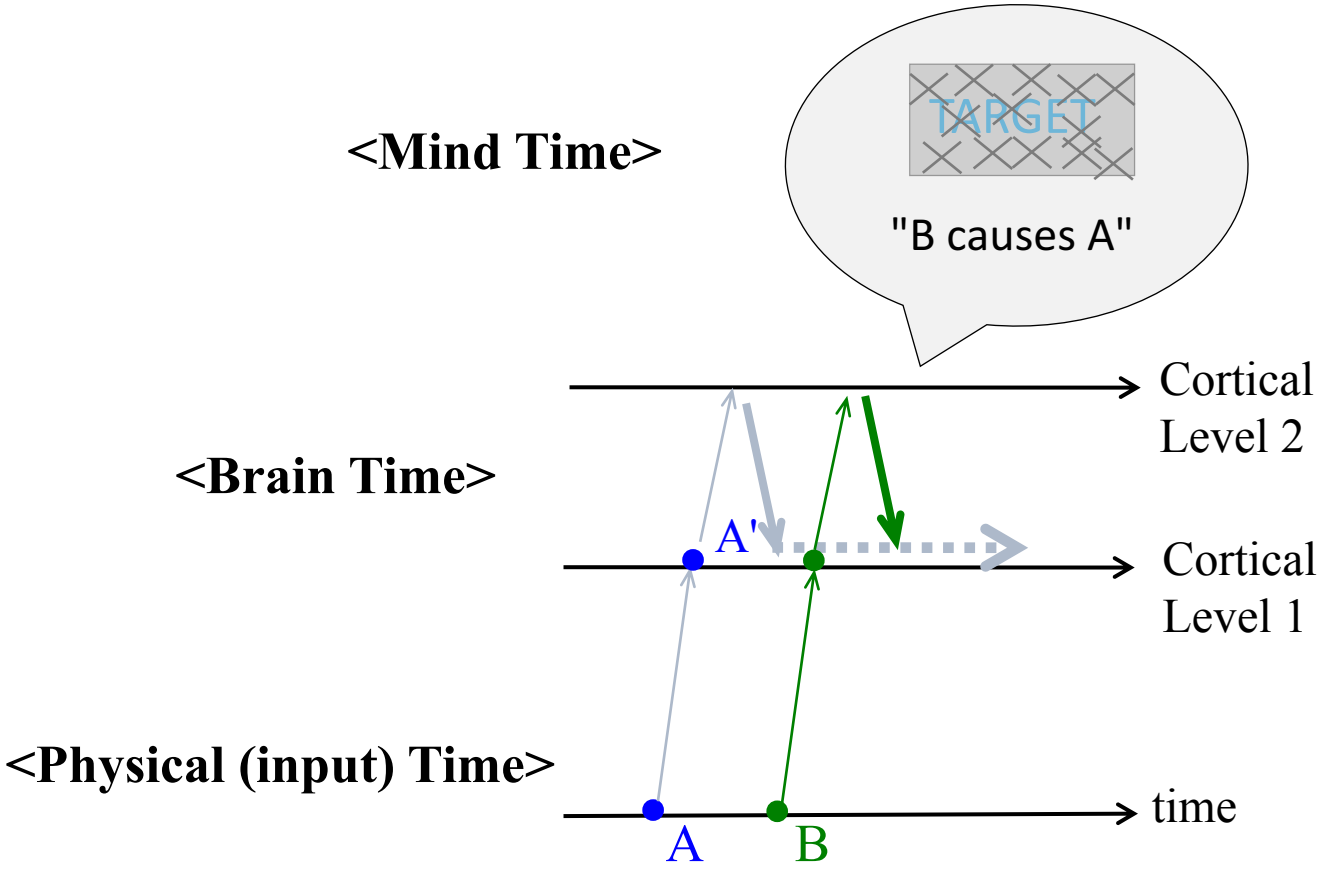
2. The feeling of free choice is ***a matter of content*** in perception/ cognition. It should ***not*** be confused with the deterministic nature of the neural correlates. Mind Time / Brain Time.

   Koeler's psychophysical (1st order) isomorphism.
   Libet ('04) ; Nishida & Johnston ('02);

   eg. (1) color perception and the "color" of neurons
   (2) pain and somatosensory cortex ("homunculus")

A neural process may causally determine that a given action is ***felt*** **voluntary** or not (as the cognitive content), whereas the neural process remains **entirely deterministic**.

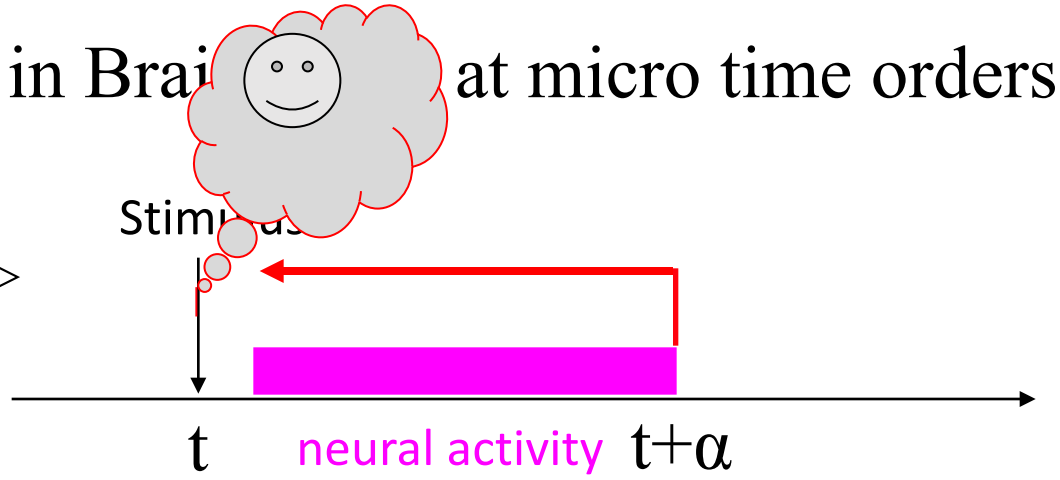
"Reentry," Neural time and perceived time



Physical sequence of the neural events (A→B) should ***not*** constrain perceived sequence ("B→A") as the ***content*** of perception.

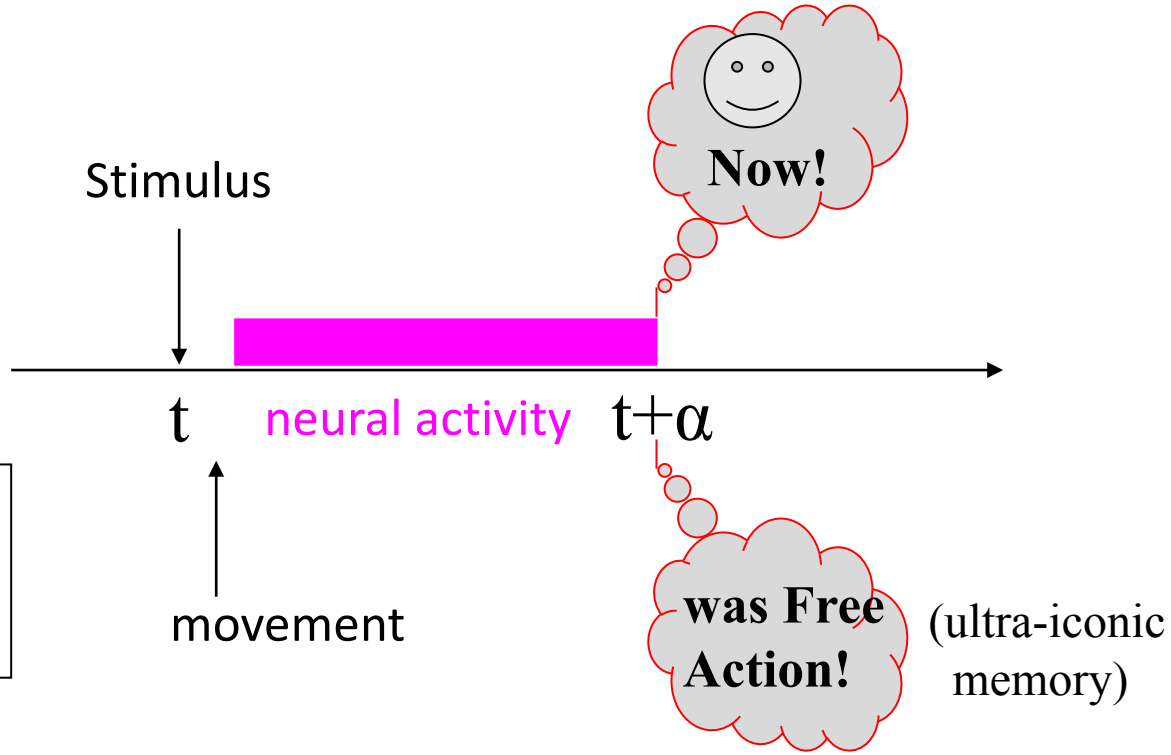
What is happening in Brai at micro time orders

＜Backward referral (Libet＞



(It looks paradoxical only because brain/mind times are plotted on one axis.)

＜Neural／perceptual time＞



1st orde isomorphism (W. Kohler) should be abandoned. No need to feel obliged to the neural/physiological sequences.

Three reasons why neural reductionism won't eliminate the sense of agency

(Shimojo fpsyg'14)

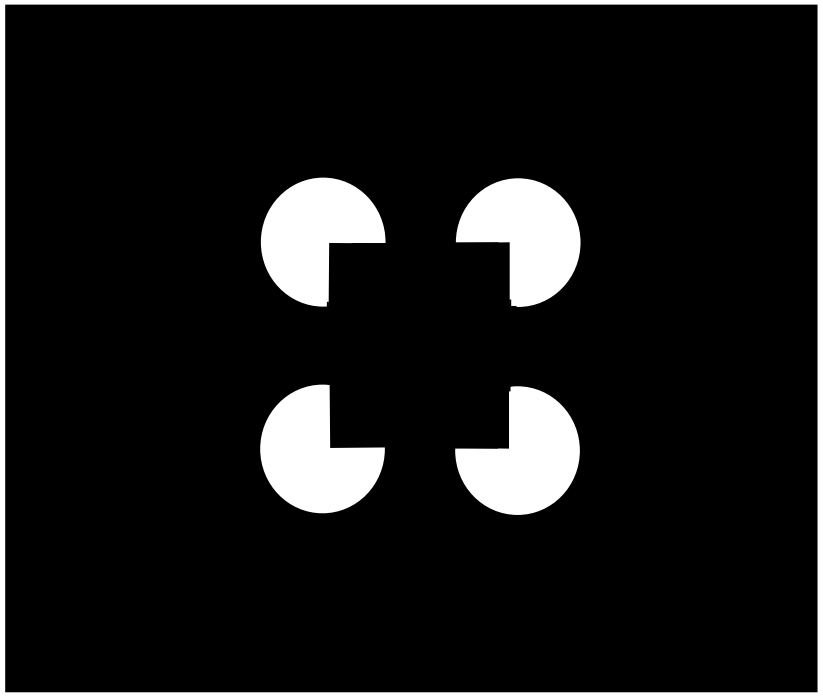
1. The feeling of free choice may live in the ***postdictive*** process, not in the predictive process.

2. The feeling of free choice is ***a matter of content*** in perception/ cognition. It should ***not*** be confused with the the deterministic nature of the neural correlates. Mind Time / Brain Time.

3. A feeling of free choice is very much like a perceptual illusion, in that it occurs in most NTs, and it will ***not*** be eliminated by objective knowledge.

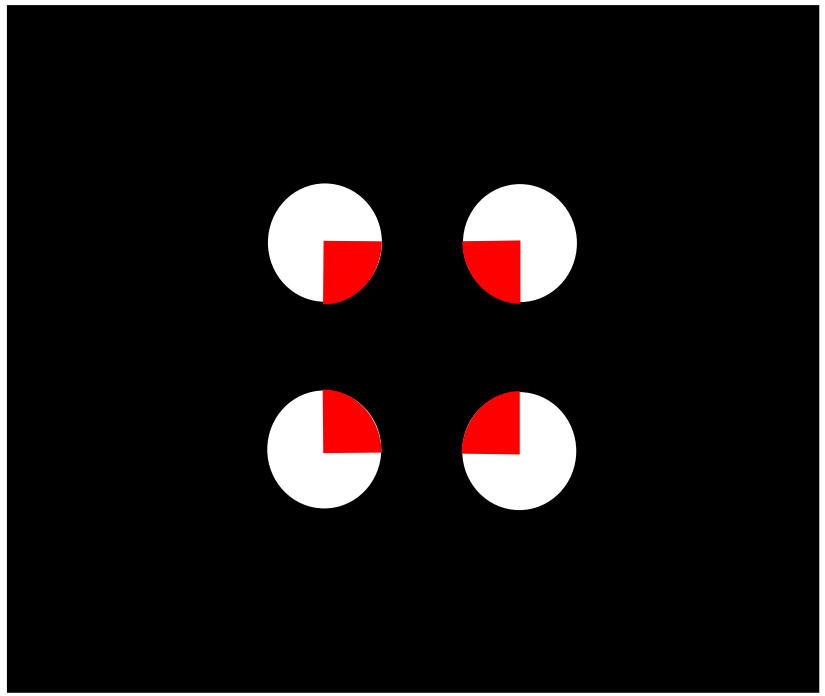
   eg. color perception, tooth pain.

Illusory Contour



The contour percept won’t disappear when knowing it does not physically exist.

Neon-color spreading & surface filling-in



The impression of transparency cannot be entirely eliminated by ”correct” knowledge.

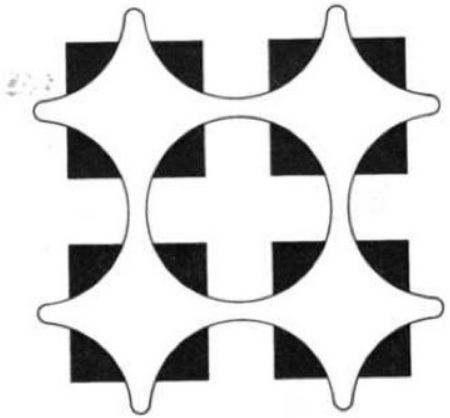


Figure 1.5a Amodal completion creates four black squares (see Figure 1.5d).

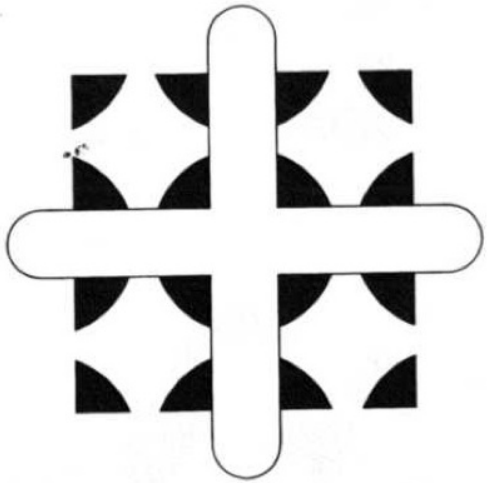


Figure 1.5b With the same black elements (see Figures 1.5c and 1.5e) only circular patterns are seen.

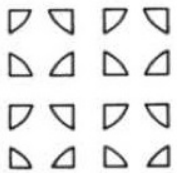


Figure 1.5c

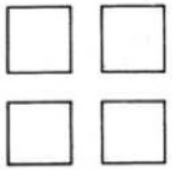
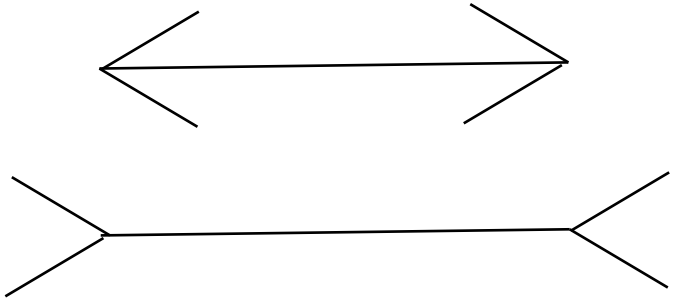


Figure 1.5d

Figure 1.5e

Perceptual organization (grouping) is automatic, not affected by cognitive knowledge(Kanizsa, 1979).



Not modified by knowledge.

Three requirements of “valid illusion”

1. (Almost) all NeuroTypicals experience it ***robustly***.
2. It reflects ***normal***, not abnormal, adaptive brain functions.
3. Resistant ***against*** top-down ***knowledge***.

The sense of agency satisfies these three conditions.

Wegner, D.M. : "Free will as a cognitive illusion"

(Wegner & Wheatley, '99; Wegner, '02)

People can experience conscious will quite ***independent*** of any actual causal connection between their thoughts and actions. The impression that a thought has caused an action rests on an ***apparent*** causal sequence.

(1) priority, (2) consistency, (3) exclusivity.

* Similar, but we further argue:
  - Postdictive construction of the brain; ***an "authentic" illusion***.
  - (1), (2), (3) above can well be postdictive constructs.

* ***Cognition can easily become perception***.
  - Free will as a predominant fiction in social system.
  - Lung cancer is immediately perceived X-ray experts.
  → Same for consciousness?

It boils down to : how to understand the "left-over" sense of freedom.

→ Opacity of human knowledge

1. Difference in perspectives -A knows it, but B does not. (*cf.* H.G. Frankfurt's "over determination" against free will.)
2. Implicit mind: not accessible to self knowledge (or meta cognition).
3. Subconscious intention(?) - no need to be consciously aware (at the moment), only has to be recognized as such (intention) postdictively.
4. Different levels of description (micro/individual/social).
5. Asymmetry among past-present-future
   a) qualitative differences in opacity.
   b) present as the distinctive point of perspective).

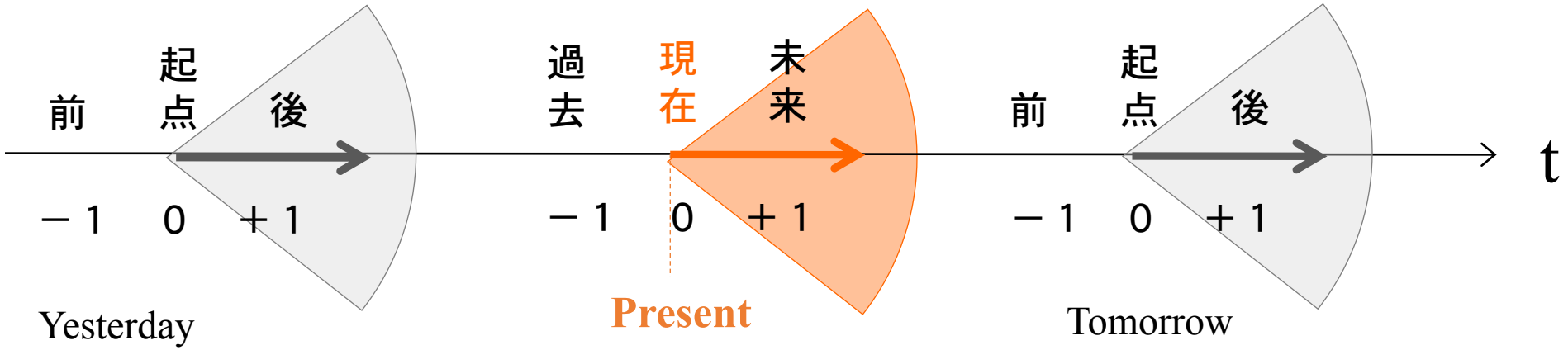
Opacity of human knowledge, and the phenomenology of time

＊ We often do not know facts (already occurred) or information (available there).

＊ Also, possible to know at some moment (later).

cf. The Unreality of Time by J. McTaggart (1908)

A/B/C series; The distinctive point of ”present” disappears in B series.



Accepting “present” as a distinctive point of temporal perspective, then any (physical, or mathematical) theories not having this may be irrelevant.

Fatalism, and the Opacity of human knowledge

"London Airport" (revised)

You're about to take a flight. If the aircraft is in fate to fall and to crash, then no point to pray. If not, then no need to pray or anything.

"A retroactive prayer?" (revised)

Your son might have been on the flight which crashed last night. Any point to pray for him now? If he was there and dead already, then no point to pray. If he was not and safe, then also no point to pray.

(revised from M. Irifuji, 2002)

← There is no paradox or issue, assuming the opacity of human knowledge.
It saves "value of prayer" from the fatalism (?).
Also, this structure(religious prayer/fatalism) is analogous to free will/neural determinism.

Passive vs. Active Decision Making

What makes this phenomenological difference? - an entirely missed point.

- Nature of tasks - tautological.
- Neural correlates (different brain loci) - somewhat tautological.
- Dependency on on-line sensory signals? – maybe not.
- Active orienting - maybe. (← gaze cascade)
- Attribution - related? (← "choice blindness" Johansson, et al.)

Internal state of the brain? - promising for a crossmodal illusion.
It may be necessary, but not sufficient.

- Prediction vs. Postdiction - promising?

SUMMARY

1. Introduction - Neural determinism as the central dogma.
2. Contents of Neural determinism/reductionism - RP, etc.
3. Three reasons why neural reductionism won't eliminate the sense of agency:
   1) Free will lies in the postdictive process.
   2) Content of perception/cognition can be different from underlying neural sequences.
   3) Free will as an authentic illusion.
4. Opacity of human knowledge (= existentialistic aspect of life), and resolution of the paradox.
5. Implications to human neuroscience - passive vs. active decision.