A DFT architecture for an intentional agent

Gregor Schöner Institute for Neural Computation <u>dynamicfieldtheory.org</u>

Doctoral dissertation of Jan Tekülve

[Tekülve, Schöner, IEEE Trans Cog Dev Sys 2022; Tekülve, Schöner Cog Science, in press (2024)]

Do the concepts of DFT potentially reach all processes of the mind?

- so that DFT would provide a neural foundation for understanding the mind?
- in the second second

Do the concepts of DFT potentially reach all processes of the mind?

- borrowing terms from the philosophy of mind to explore this question
- "intentionality" in two "directions of fit" to sample qualitatively different form of mental and motor acts
- "psychological modes" to sample the mind from the sensory-motor to goalachievement and knowledge

Intentionality

- Intentionality = the capacity of organisms and their nervous systems to generate mental states that are about things in the world
 - *things* may include an organism's own body
 - things may ultimately also includes the nervous system's own states

Two directions of fit of intentional states (Searle)

world-to-mind: the world must match the intentional state to fulfill that state's condition-of-satisfaction (CoS) => the "motor" flavor of intentionality

mind-to-world: the intentional state must match the state of the world to fulfill the CoS => "perceptual" flavor of intentionality

Six psychological modes of intentional states (Searle)

intention-in-action

prior intention

desire



The six modes reflect the sensory-motor grounding of cognition world-to-mind

📕 motor control

📕 action plans, decisions, sequences

goals, motivations, emotions

mind-to-world

- attention, active perception, working memory
- scene and event memory
- back-ground knowledge, learning from experience, communication



Six psychological modes of intentional states (Searle)



📕 world-to-mind

- intention-in-action ~ action
- 📕 prior intention ~ plans
- 📕 desire ~ goals



Six psychological modes

Hypothesis: these psychological mode reach all of the mind





lintention-in-action

📕 prior intention







reach to take up paint

reach to apply a coat of paint





memory of visual scene prior intentions search to paint search to load paint reach to apply paint move to a recalled

move to a recalled location ...



beliefs (propositional)

rules linking color concepts: which paint on which canvas generates which new color

learn these beliefs

desires

to point cubes in a particular color

Neural dynamic architecture



Mind-to-world intentionality

Perception! Is input driven ...

=> the intentional state is its own CoS

by virtue of how the state came about

=> time structure: intention/CoS co-active





The sensory surface



retinal space



Visual exploration



From perception to scene memory [memory initially empty, then sequentially built]



Mind-to-world intentionality

Perception! Is input driven ...

=> the intentional state is its own CoS

by virtue of how the state came about

=> time structure: intention/CoS co-active

Memory as an intentional state is also its own CoS...

and its creation is input driven

but its activation in recall... has the opposite direction of fit ... we'll get to that

The belief sub-architecture



Learning a new belief



Learn a new belief

[while exploring: applying blue paint to yellow cube]





Scene



Mind-to-world intentionality

Memory as an intentional state is also its own CoS...

and its creation is input driven

but its activation in recall... has the opposite direction of fit ... we'll get to that

Beliefs as intentional states are their own CoS...

- and their creation is input driven...
- but their (opportunistic) activation... has the opposite direction of fit

World-to-mind intentionality

intentional state persists UNTIL its CoS is activated

then both are deactivated











Recall from memory

- is a "motor" intention... (a world-to-mind intention)
- as it is aimed at achieving a particular state of the mind (which is part of the world)



chaining to organize the paint behavior



Spray



Spray

uses transient detection to detect the change of color



-0.5

0

-1

-2

-3

0.5

0

-0.5

Recall and drive



Recall and drive



Recall-drive-search

[based on a desire and an activated belief, looking for a tall pink object, which is in memory]



Recall a belief

[triggered by a desire and objects in scene memory]



The desire sub-architecture



Achieving a desire



Instantiating mind-to-world intentions are world-to-mind intentions!

here: cued recall from memory

- other examples: cued or opportunistic activation of knowledge (beliefs), visual search
- => time structure: intention active until its CoS becomes active, then de-activated

Instantiating mind-to-world intentions are world-to-mind intentions!

most thinking is of that nature !

so the time-structure and mechanism for autonomy of the world-to-mind intentions is central to cognition!

Achieving a desire



Conventional NN

e.g. in DNN

- support input-driven processes: perception, categorization
- do not have world-to-mind forms of intentionality (other than by non-neural means)
- => they do not have a meaningful way to autonomously instantiate mind-to-world intentions (other than by non-neural means)

What if the CoS does NOT happen?

two cases...

- a) nothing happens in the CoS field/subspace
- b) something happens in the CoS field/ subspace that differs from the prediction

a) nothing happens in the COS field

example: change detection

the "same" response



[Johnson, et al. 2009]

a) nothing happens in the COS field

- "same" response as the default state
- that arises if there is no "different" response from change detection



Feature Dimension

[Johnson, et al. 2009]

a) nothing happens in the COS field

 "same" is the Condition of Dissatisfaction (CoD) of the change detection task

generally: CoD as a time out of a "clock"

that is started by the onset of the intention..



[Johnson, et al. 2009]

b) something happens in the CoS field that differs from the prediction



[Grieben, Schöner, CogSci 2021]

Conclusions

- I) Insight that the "world-to-mind" direction of fit is fundamental of acting and thinking ... and is realized in DFT through the CoS concept
- 2) DFT does reach all psychological modes
- 3) DFT architectures scale..
 - stability => robustness
 - non-synesthesia principle

Outlook: beyond action

planning...

an implied hierarchy...

goals/desires

goal dynamics... opportunistic activation of goals

🛑 goal habituation

motivation

Frontier: world-to-mind intentions persisting over long times

as evidenced in cognitive control...

- perhaps only intermittently active, then reactivated?
- timers at different time scales?