

Eric Werner  
Oxford Advanced Research Foundation  
<https://www.oarf.com>  
<https://www.ericwerner.com>  
©Eric Werner 2023, All rights reserved.

Rough Draft V3 October 10, 2023

# Social Information in Artificial Social Intelligence

Social information can make Large Language Models  
and Large Behavioral Models cooperative<sup>1</sup>



Colors show meta-interaction protocol in a developed Fish Brain

---

<sup>1</sup>©Eric Werner, All rights reserved 2023.

## **The Minimal Social Brain**

What would a minimal model of an agent's neural network-mind-brain such as human be like? What are its minimal characteristics? What is the global architecture? How would we map the agent's neural network to our abstract model? What is the minimal complexity of the architecture that captures the essence of what it is to be an intelligent agent in the world?

Can a purely linguistic interaction between a human and a dialogue-prompter (computational device) capture the mental capacities, the network, the network state and information states of the human agent's brain?

## **Short history of communication theory**

There is an inherent entropy (ambiguity) in the semantics and pragmatics of any linguistic sentence. Hence, a purely linguistic interaction would inevitably generate an entropic model of the agent's brain-network. However, the model may be functionally equivalent to the agent's brain-network in many situations of use. Wittgenstein's notion of meaning as use (see 'Blue and Brown Books', 'Philosophical Investigations') (Wittgenstein 1958) is relevant here since it suggests functional equivalence is sufficient for human communication.

John von Neumann mathematical theory of games suggests that there is a minimal architecture for agents in games and, more generally, in the economic world (von Neumann J. 1947). In (Werner 2023 forthcoming) I investigate the logic of information in games, specifically the logic of information states in agents and how

they evolve in time. Later I extended that work to investigate the logic of intentions and social action in agents with limited information about their world. This extended von Neumann's concept of information-set to include agent strategic states (that give a model of the agent's intentional states). This permitted the formalization and understanding of multi-agent cooperation by way of interacting intentional states of agents(Werner 1988).

Wittgenstein realized the inadequacy of a purely information based theory of language and meaning. In his *Philosophical Investigations*(Wittgenstein 1958) and the *Blue and Brown Books*(Wittgenstein 1958) he viewed meaning a given by use, such as in a language game. Searle followed with his work on speech acts. Later Habermas (Habermas 2015, Habermas 2015) used these concepts for his work on social communication. However, all of these theories of language and social action were based on informal vague notions of meaning (semantics and pragmatics). Chomsky theories of syntax were just that with no semantics, no pragmatics, no theory of communication.

In (Werner 1999) (Werner 1991) (Werner 1996) (Werner 1988) I generalized information-based communication theory going beyond (Shannon 1949), (von Neumann J. 1947), logical foundationalism (e.g., (Whitehead and Russell 1950) early Wittgenstein's *Tractatus* (Wittgenstein 1969) to include intention-strategic-based communication and cooperation theory. On my view communication involves linguistic intentional states formalized as linguistic strategic states linked to the world via semantics and pragmatic meaning.

Moreover, I showed the intimate relationship between states of information and what an agent can do and what the agent can intend in a complex social world of other agents(Werner 1991). This work unified the logics of information (world-state-information), intention (strategic information) and ability (the logic of can and social ability). Add to that von Neumann's formalization of utility and you get the foundational bedrock of the minimal architecture of mental states of communicating social agents whether they be human, animal or robotic.

How does this relate to artificial intelligence (AI) and ChatGPT, general AI (GAI) and autonomous robots (self driving cars, social robots, ethical robots)?

## **The social brain**

How is the architecture of social agents and social information relate to the architecture of the human brain, the connectome, cerebral cortex, visual cortex, Wernicke area, Broca area, and the cerebellum?

Where is social information and intentional capacity located in the human brain?

## **The architecture of a social brain**

I would like to discuss the overall architecture of the social brain . The fundamental question is: What makes a society possible at all? What is the minimal architecture of the brain to enable an agent to be social?

Communication by language is obviously central, but it requires more. It requires a particular mental capability, and the parts of the agents that have social capacities. The capacity of the social brain is the ability to represent intentions, not only the intentions of self, but the intentions of others. Through the interaction of the representation of intentions of others, and the representation of intentions of the self, it makes coordination and cooperation possible.

Intentions can be about actions, but they can also be about linguistic actions. Coordinated speech is one of the fundamental processes for achieving coordinated intentions for multi-agent intentional states. Intentional states are fundamentally different than knowledge about the world which we call information states. Intentional states require informational states because the agent needs to know about the world in order to be able to act. The agents abilities are direct result of the information the agent has about the world.

The less information that the agent has about the world, the fewer the things that the agent can do. Thus the logic of ability is directly related to the logic of information. The logic of ability underpins the logic of intention.

## **Information, Intention and Utility**

There is one more component and that is value or utility. Utility drives the formation of intentions of the agent. Therefore, to get a minimal social agent we need at least three components: **Intention S**, **information I**, and **utility** or **value** (Werner 1988)r.

I called is three components, the **Representational Capacity** of the agent  $R = (S, I, V)$ .

A fourth element may be called planning and logical inference i.e., **Planing-Logic** PL.

It may be that each one of these components ( S, I, V, PL ) will require different types of training for a ChatGPT-like Large Language Model (LLM) or Large Behavior Model (LBM).

They correspond to different areas of the human brain (e.g., the visual cortex, Wernicke area (language and visual understanding), Broca area (speech generation), and more, of course. The connectome located in the White Matter of the brain links the different areas. Its role is fundamental for brain coordinated processing (Geschwind 1974). For a nice overview of the connectome research see (Catani, Sandrone et al. 2015)

## **Artificial Social Intelligence**

Social information is different than information about the state of the world. Intentional strategic states are implemented through social information representations. The whole system of human, linguistic communication is an example of functional social information exchange. An intentional state is the prime example of social information.

## **The social genome and the social brain**

We can view the genome as a social structure, which contains both the two parental genomes. There's an interaction between the parental genomes, by means of a protocol, a meta-protocol that determines which genome is in control at which time in the development of the embryo. This enables the development of the brain to also reflect both of the parents and their social capacities. It explains the fact that the thinking of different children, maybe reflect that a one parent or that of another at different stages of development. The brain is sectioned into different areas that reflect the control of one parents genome or that of the other just as the body is sectioned into different areas as it is being controlled by one parental genome or the other parental genome.

## **Social Information**

Thus, the very development of the brain is controlled by social information processes see my “Brain meta genomics: Genome mind mapping: Network protocols partition the developing brain” (Werner 2023).

The social information is embedded at all levels of the human body. It is fundamental to the interaction between parental genomes. It is embedded at the level of cell interactions, through cell-cell communication. It is embedded in the brain in different areas of the brain that specialize in various processes. Many areas of the brain process social information such as intentions and values, utilities or emotions. Entire communication system of the brain from output to input from speech generation to speech understanding is fundamentally a social information processing system. The human brain is fundamentally social in nature.

If we are going to create cooperative LLMs or LBMs then their architecture has to reflect the architecture of the human social brain. These artificial intelligent social systems, will have to process social information much like human brains process social information.

## Bibliography

- Catani, M., S. Sandrone and A. Vesalius (2015). *Brain renaissance : from Vesalius to contemporary neuroscience*. New York, NY, Oxford University Press New York, NY.
- Geschwind, N. (1974). *Selected Papers on Language and the Brain*. Dordrecht, Springer Netherlands Dordrecht.
- Habermas, J. ê. (2015). *The theory of communicative action : 1. Reason and the rationalization of society*. [Place of publication not identified], Polity Press [Place of publication not identified].
- Habermas, J. ê. (2015). *The theory of communicative action : 2. Lifeworld and systems, a critique of functionalist reason*. [Place of publication not identified], Polity Press [Place of publication not identified].
- Shannon, C., Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, University of Illinois Press.
- von Neumann J., M. O. (1947). *The Theory of Games and Economic Behavior*. Princeton, NJ, Princeton University Press.
- Werner, E. (1988). "Toward a theory of communication and cooperation for multiagent planning." *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Second Conference*: 129-143.
- Werner, E. (1991). *A Unified View of Information, Intention and Ability*. *Decentralized AI*. Y. M. Demazeau, J.P., Elsevier Science Publishers. **2**.
- Werner, E. (1996). *Logical foundations of distributed artificial intelligence*. *Foundations of Distributed Artificial Intelligence*. G. O'Hare, and N. Jennings, Wiley.
- Werner, E. (1999). *The Ontogeny of the Social Self. Towards a Formal Computational Theory*. *Human Cognition and Social Agent Technology*. K. Dautenhahn, John Benjamins: 263-300.
- Werner, E. (2023). *Brain meta genomics Genome mind mapping: Network protocols partition the developing brain*. *Internet of Life*. To be published.
- Whitehead, A. N. and B. Russell (1950). *Principia mathematica*. Cambridge, Cambridge University Press Cambridge.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Oxford, Blackwell.
- Wittgenstein, L. (1958). *Philosophical Investigations*. Oxford, Blackwell.

Wittgenstein, L. (1969). Schriften. 1 : Tractatus logico- philosophicus, Tagebücher 1914-1916, Philosophische Untersuchungen.